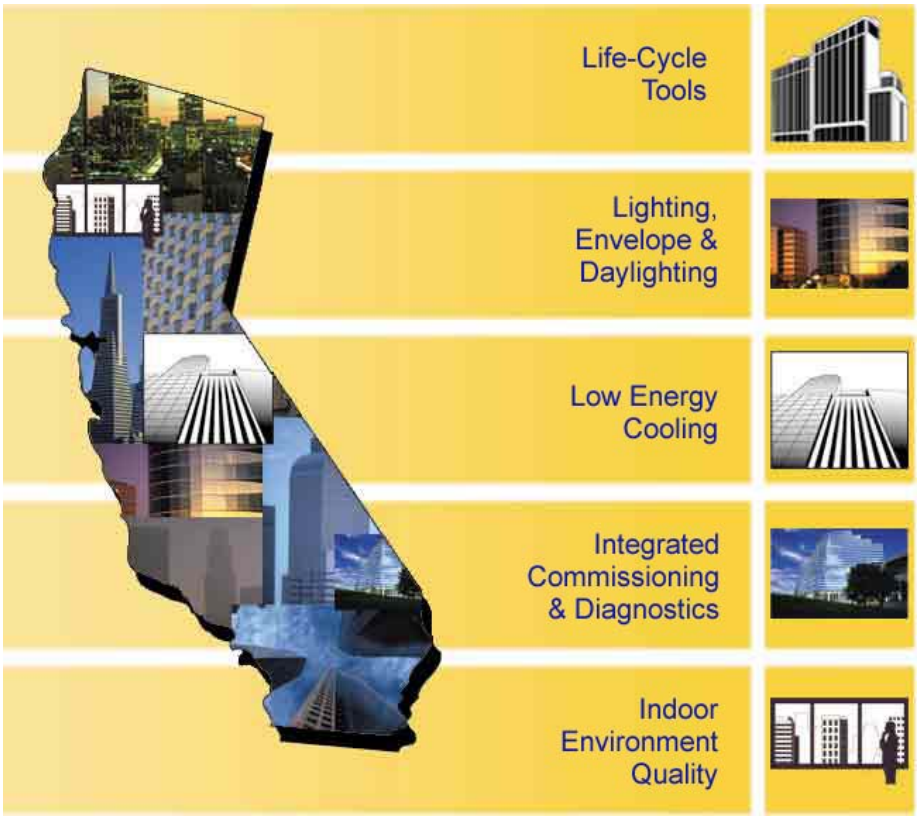
CALIFORNIA
ENERGY
COMMISSION

Non-Intrusive Load Monitors (NILMs) used for Equipment Monitoring and Fault Detection



TECHNICAL REPORT

October 2003
500-03-097-A16

Gray Davis, *Governor*

CALIFORNIA ENERGY COMMISSION

Prepared By:
Buildings Technologies Department
Lawrence Berkeley National Laboratory

Steve Selkowitz
B90R3110
1 Cyclotron Road
E. O. Lawrence Berkeley National Laboratory
Berkeley, CA 94720

CEC Contract No. 400-99-012

Prepared For:
Martha Brook,
Contract Manager

Nancy Jenkins,
PIER Buildings Program Manager

Terry Surles,
PIER Program Director

Robert L. Therkelsen
Executive Director

Acknowledgements

In a program of this magnitude there are many people who contributed to its success. We owe the many staff members, faculty, and students of the different institutions our thanks for the superb work and long hours they contributed. All of their names may not appear in this report, but their efforts are visible in the many papers, reports, presentations, and thesis that were the major output of this program.

The EETD leadership provided support in many ways. We thank Mark Levine, Marcy Beck, and Nancy Padgett. Members of the Communications Department of EETD helped in preparing reports, presentations, handouts, and brochures. The help of Allan Chen, Julia Turner, Anthony Ma, Steve Goodman, Sondra Jarvis, and Ted Gartner is acknowledged.

Special thanks are given to the support staff from the Buildings Technologies Program at LBNL: JeShana Dawson, Rhoda Williams, Denise Iles, Catherine Ross, Pat Ross, and Danny Fuller. Norman Bourassa performed a wide range of duties, from original research to tracking deliverables.

We thank the following members of the Program Advisory Committee (PAC) for their advice and support. In a program designed to deal with real world problems their ideas were vital. The PAC members are:

Larsson, Nils	C2000 Canada
Stein, Jay	E-Source
Wagus, Carl	Am. Architectural Manufs. Assoc.
Lewis, Malcolm	Constructive Technologies
Bernheim, Anthony	SMWM Architects
MacLeamy, Patrick	HOK
Mix, Jerry	Wattstopper
Waldman, Jed	CA Dept of Health Services
Bocchicchio, Mike	UC Office of the President
Prindle, Bill	Alliance to Save Energy
Sachs, Harvey	ACEEE
Browning, Bill	Rocky Mountain Institute
Lupinacci, Jean	U.S. EPA
Goldstein, Dave	Natural Resources Defense Council
Smothers, Fred	Smother & Associates
Benney, Jim	NFRC Director of Education
Stewart, RK	Gensler Assoc
Angyal, Chuck	San Diego Gas & Electric
Ervin, Christine	US Green Buildings Council
Ginsberg, Mark	US Department of Energy
Higgins, Cathy	New Buildings Institute

Finally, we acknowledge the support and contributions of the PIER Contract Manager, Martha Brook, and the Buildings Program team under the leadership of Nancy Jenkins.

Preface

The Public Interest Energy Research (PIER) Program supports public interest energy research and development that will help improve the quality of life in California by bringing environmentally safe, affordable, and reliable energy services and products to the marketplace.

The Program's final report and its attachments are intended to provide a complete record of the objectives, methods, findings and accomplishments of the High Performance Commercial Building Systems (HPCBS) Program. This Commercial Building Energy Benchmarking attachment provides supplemental information to the final report (Commission publication # 500-03-097-A2). The reports, and particularly the attachments, are highly applicable to architects, designers, contractors, building owners and operators, manufacturers, researchers, and the energy efficiency community.

This document is the sixteenth of 22 technical attachments to the final report, and consists of research reports:

- Monitoring HVAC Equipment Electrical Loads from a Centralized Location – Methods and Field Test Results (E5P2.2T3a)
- Detection of HVAC Faults via Electrical Load Monitoring (E5P2.2T3b)
- Demonstration of Fault Detection and Diagnosis Methods for Air-Handling Units (ASHRAE 1020-RP) (E5P2.2T3c)
- Power Signature Analysis (E5P2.2T3d)

The Buildings Program Area within the Public Interest Energy Research (PIER) Program produced this document as part of a multi-project programmatic contract (#400-99-012). The Buildings Program includes new and existing buildings in both the residential and the nonresidential sectors. The program seeks to decrease building energy use through research that will develop or improve energy-efficient technologies, strategies, tools, and building performance evaluation methods.

For the final report, other attachments or reports produced within this contract, or to obtain more information on the PIER Program, please visit http://www.energy.ca.gov/pier/buildings or contact the Commission's Publications Unit at 916-654-5200. The reports and attachments are also available at the HPCBS website: http://buildings.lbl.gov/hpcbs/.

Abstract

Monitoring HVAC Equipment Electrical Loads from a Centralized Location – Methods and Field Test Results: Documenting Meter Tests at the Iowa Energy Center

This report presents the field-test results of the steady-state non-intrusive load monitoring (SS-NILM) system developed at M.I.T., and its suitability to load monitoring and fault detection and diagnosis in a small commercial building, specifically a KFC Restaurant in Norwell,Massachusetts. The results from the NILM system were validated using an independent and "traditional" multi-channel end-use power-metering system installed at the site.

The organization of this report is as follows: First a general description of the NILM system developed at M.I.T., hardware and software, is given in chapter two, followed by a description of the test sites selected for this project in chapter three. The site description includes the equipment connected to the electrical panel monitored by the NILM system as well the parallel power-metering system, installed to validate the results obtained from the NILM system. A discussion and comparison of the results obtained from the NILM and the parallel monitoring systems, as well as the modifications made to the NILM software components based on these results are presented in chapter four of the report. Conclusions and recommendations for possible future work are discussed and presented in the final chapter of the report.

Detection and Diagnosis of HVAC Faults via Electrical Load Monitoring

Detection and diagnosis of faults (FDD) in HVAC equipment have typically relied on measurements of variables available to a control system, including temperatures, flows, pressures, and actuator control signals. Electrical power at the level of a fan, pump, or chiller has been generally ignored because power meters are rarely installed at individual loads. This paper presents two techniques for using electrical power data for detecting and diagnosing a number of faults in air-handling units. The results from the two techniques are compared and the situation for which each is applicable is assessed.

One technique relies on gray-box correlations of electrical power with such exogenous variables as airflow or motor speed. This technique has been implemented with short-term average electrical power measured by dedicated submeters. With somewhat reduced resolution, it has also been implemented with a high-speed, centralized power meter that provides component specific power information via analysis of the step changes in power that occur when a given device turns on or off. This technique was developed to detect and diagnose a limited number of air handler faults and is shown to work well with data taken from a test building. A detailed evaluation of the method is presented in the companion paper, which documents the results of a series of semiblind tests.

The second technique relies on physical models of the electromechanical dynamics that occur during the short-time seconds after a motor is turned on. This technique has been demonstrated

with submetered data for a pump and for a fan. Tests showed that several faults could be successfully detected from motor startup data alone. While the method relies solely on generally stable and accurate voltage and current sensors, thereby avoiding problems with flow and temperature sensors used in other fault detection methods, it requires electrical data taken directly at the motor, downstream of variable-speed drives, where current sensors would not be installed for control or load-monitoring purposes.

Demonstration of Fault Detection and Diagnosis Methods for Air-Handling Units (ASHRAE 1020-RP)

Results are presented from controlled field tests of two methods for detecting and diagnosing faults in HVAC equipment. The tests were conducted in a unique research building that featured two air-handling units serving matched sets of unoccupied rooms with adjustable internal loads. Tests were also conducted on a third air handler in the same building, serving areas used for instruction and by building staff. One of the two fault detection and diagnosis (FDD) methods used first-principles-based models of system components. The data used by this approach were obtained from sensors typically installed for control purposes. The second method was based on semiempirical correlations of submetered electrical power with flow rates or process control signals.

Faults were introduced into the air-mixing, filter-coil, and fan sections of each of the three air-handling units. In the matched air-handling units, faults were implemented over three blind test periods (summer, winter, and spring operating conditions). In each test period, the precise timing of the implementation of the fault conditions was unknown to the researchers. The faults were, however, selected from an agreed set of conditions and magnitudes, established for each season. This was necessary to ensure that at least some magnitudes of the faults could be detected by the FDD methods during the limited test period. Six faults were used for a single summer test period involving the third air-handling unit. These fault conditions were completely unknown to the researchers and the test period was truly blind.

The two FDD methods were evaluated on the basis of their sensitivity, robustness, the number of sensors required, and ease of implementation. Both methods detected nearly all of the faults in the two matched air-handling units but fewer of the unknown faults in the third air-handling unit. Fault diagnosis was more difficult than detection. The first-principles-based method misdiagnosed several faults. The electrical power correlation method demonstrated greater success in diagnosis, although the limited number of faults addressed in the tests contributed to this success.

The first-principles-based models require a larger number of sensors than the electrical power correlation models, although the latter method requires power meters that are not typically installed. The first-principles-based models require training data for each subsystem model to tune the respective parameters so that the model predictions more precisely represent the target system. This is obtained by an open-loop test procedure. The electrical power correlation method uses polynomial models generated from data collected from “normal” system operation, under closed-loop control.

Power Signature Analysis -Advanced Nonintrusive Monitoring of Electric Loads

The Nonintrusive Load Monitor (NILM) can determine the operating schedule of electrical loads in a target system from measurements made at a centralized location, such as the electric utility service entry. In contrast to other systems, the NILM reduces sensor cost by using relatively few sensors. The NILM disaggregates and reports the operation of individual electrical loads like lights and motors using only measurements of the voltage and aggregate current at the utility service point of a building. It can also identify the operation of electromechanical devices in other kinds of power distribution networks. For example, the NILM can determine the load schedule in an aircraft from measurements made only at the generator, or in an automobile from measurements made at the alternator/battery block. The NILM can distinguish loads even when many are operating at one time. The NILM is an ideal platform for extracting useful information about any system that uses electromechanical devices. The NILM has a low installation cost and high reliability because it uses a bare minimum of sensors. It is possible to use modern state and parameter estimation algorithms to verify remotely the health of electromechanical loads by using a NILM to associate measured waveforms with the operation of individual loads. The NILM can also monitor the operation of the electrical distribution system itself, identifying situations where two or more otherwise healthy loads interfere with each other's operation through voltage waveform distortion or power quality problems. Strategies for nonintrusive monitoring have developed over the last twenty years. Advances in computing technology make a new wealth of computational tools useful in practical, field-based NILM systems. This paper reviews techniques for high performance nonintrusive load and diagnostic monitoring and illustrates key points with results from field tests.

California Energy Commission
Public Interest Energy Research Program

HPCBS

High Performance Commercial Building Systems

Monitoring HVAC Equipment Electrical Loads from a Centralized Location Methods and Field Test Results (ASHRAE AC-02-12-4)

Element 5 - Integrated Commissioning and Diagnostics
Project 2.2 - Monitoring and Commissioning of Existing Buildings

Dong Luo
United Technologies Corporation
Leslie K. Norford and
Massachusetts Institute of Technology
Steven R. Shaw
Montana State University

ASHRAE Transactions 2002
Volume 108, Part 1

pier
BERKELEY LAB
United States Department of Energy
Science, Security and Energy: Powering the 21st Century

Acknowledgement

This work was supported by the California Energy Commission, Public Interest Energy Research Program, under Contract No. 400-99-012 and by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

AC-02-12-4

Monitoring HVAC Equipment Electrical Loads from a Centralized Location—Methods and Field Test Results

Dong Luo, Ph.D.
Student Member ASHRAE

Leslie K. Norford, Ph.D.
Member ASHRAE

Steven R. Shaw, Ph.D.

Steven B. Leeb, Ph.D.

ABSTRACT

This paper reports recent work to determine useful information about component-level HVAC electrical loads—not from submeters, which are accurate, rarely installed, and relatively expensive, but instead from one or more centralized locations in a building's electrical distribution system. The work includes laboratory tests with real-building data and field tests made with low-cost hardware capable of the rapid sampling needed for load disaggregation. Results indicate that building electrical signals are often quite complex, that individual loads can indeed be detected with reasonable reliability, that more work is required to automate the process of tuning the detection algorithm, and that there are benefits to analyzing turn-on/turn-off events at multiple sampling rates to minimize trade-offs between detection sensitivity and false alarms.

INTRODUCTION

Accurate and affordable information about HVAC electrical loads is of value to many individuals and organizations involved in providing HVAC services: facility managers, who would like to minimize operating costs and the costs and down-time associated with repairs; electric utilities and service providers, who need accurate load models to most economically generate, transmit, and distribute power; and energy service companies and building owners, who would like inexpensive means to verify savings from energy-efficient improvements. Electrical-power information can also be used for power-quality monitoring and for analyzing loads other than HVAC, including lights and process equipment.

The demands for both reasonable accuracy, suitable for the task at hand, and moderate cost, in keeping with achievable benefits, are not easily met. The required accuracy varies with specific monitoring goals, graduated as follows:

1. detection of on-off switching events, to determine whether equipment is operating in accordance with the expected schedule and whether it responds to a control signal;
2. measurement of power magnitudes at the time of on-off switches and throughout the operating cycle, to compute energy consumption; and
3. measurement of changes in power or changes in the frequency of operation of equipment, both relative to normal operation, as a basis of detecting faulty operation.

The first level can be easily achieved with current transducers, now in limited use in buildings to effectively provide an echo of a control signal and thereby verify that a fan or pump has turned on or off on command. This simple technology is not universally applied, suggesting some ambivalence about cost versus benefit. Going to the second and third levels requires more expensive power metering rather than binary (on-off) current information. The third level, fault detection and diagnosis, also requires development of methods to analyze changes in electrical-power data and to relate these changes to normal operating patterns.

This paper describes recent work on an electrical-load monitor capable of obtaining electrical-power data at a cost lower than submeters. This monitor has its origins in residential buildings, where it was designed to be installed in lieu of the standard revenue meter and hence did not cross the utility-customer boundary. In that sense it was not invasive and was therefore known as a non-intrusive load monitor, or NILM. The analysis method for the residential meter, detailed in Hart (1992) and now in commercial production, is based on

Dong Luo is a senior engineer with United Technologies Corporation. **Leslie K. Norford** and **Steven B. Leeb** are associate professors at the Massachusetts Institute of Technology, Cambridge, Mass. **Steven R. Shaw** is an assistant professor at Montana State University, Bozeman.

pinpointing the times at which a near-constant series of electrical power measurements changes to another near-constant series. Changes are characterized by their magnitude in real and reactive power. Changes of near-equal magnitude and opposite sign are paired to establish the operating cycles and energy consumption of individual residential appliances.

This approach faces several limitations in commercial buildings and, increasingly, in houses:

1. electrical noise generated by power electronics, which makes it difficult to establish steady-state conditions and, once steady state is defined, also generates changes in power that must be analyzed as events even though they are not of interest;
2. overlapping on-off events that may mask individual changes; and
3. time-varying electrical power demand by individual components.

The first two limitations have been overcome in two ways. First, as described in detail elsewhere (Leeb 1993; Leeb et al. 1995; Norford and Leeb 1996; Leeb and Kirtley 1996; Leeb et al. 1998), a method has been developed to measure and analyze the short-term dynamic electrical pattern associated with the start-up of a piece of equipment to aid in identifying this equipment in an environment that is either noisy or rich with information. This approach computes spectral envelopes at the fundamental and higher harmonic frequencies, detects rapid changes in these envelopes, and compares the patterns of such changes with libraries of known patterns, generated for either classes of equipment (induction motors, rapid-start lamps) or individual components (a grinder, for example). Another approach to analyzing start-up transients as a means of non-intrusive load detection is presented in Deschizeau et al. (2000).

Detection and analysis of start-up transients hold the promise of a powerful approach to fault detection, requiring only short-term, focused, and robust power analysis rather than more extended computation of changes in power consumption under known loading conditions. Recent efforts in this area rely on submetered rather than centralized power measurements and are described in Shaw et al. (2002).

Second, a less powerful, more easily implemented, and ultimately complementary effort detects changes in power levels in a manner that is similar in spirit to the original residential concept but more effective in noisy electrical environments encountered in commercial buildings. Norford and Leeb (1996) showed examples of how one aspect of this work, median filtering, can improve signal quality. Abler et al. (1998) introduced the signal processing algorithms. Work reported in this paper builds on the initial efforts of Hill (1995), provides a full discussion of material introduced in Abler et al. (1998), and presents a significant extension to cope with signals of interest that are small relative to noise. This approach has immediate application in a number of areas, including fault detection and diagnosis, and also provides a means of triggering the more computationally intensive transient-event detector.

This paper first describes an appropriate technique for detecting turn-on/turn-off events in noisy electrical power signals. The basic method precedes a number of refinements. The refined approach is then melded with an oscillation detector, to shield the on-off detector from the impact of oscillatory power signals and to analyze those oscillations as indicators of poorly tuned HVAC controllers. Next, the on-off detector is extended to operate at multiple sampling rates, found necessary in order to reliably discern cycling patterns for a two-stage reciprocating chiller in a whole-building electrical signal. Finally, the paper compares centralized power measurements with submeters, briefly describes the monitoring hardware, and offers conclusions. Most of the data used in this work were collected at the test building used for an ASHRAE-sponsored research project on fault detection in HVAC systems, RP 1020; the test building is described more fully in Norford et al. (2000, 2002).

BASIC DETECTION ALGORITHM—GENERALIZED LIKELIHOOD RATIO (GLR)

Figure 1 shows representative electrical-power data collected at the ASHRAE test building. Electrically noisy commercial buildings have led to a large number of false alarms when detecting turn-on/turn-off events via changes in steady-state power (Norford and Mabey 1992). A statistical algorithm more reliable and powerful than a simple trigger based on deviation from the mean has been developed by extending the generalized likelihood ratio (GLR) (Basseville and Nikifirov 1993).

The GLR detection algorithm calculates a decision statistic from the natural log of a ratio of probability distributions before and after a potential change in mean:

$$S_j^k(\mu_1) = \sum_{i=j}^{k} \ln \frac{P_{\mu_1}(y_i)}{P_{\mu_0}(y_i)}, \qquad (1)$$

where

y_i = sampled variable at time i;

μ_0, μ_1 = mean values of the sampled sequence before and after the event, respectively;

$P_\mu(y_i)$ = probability density function of the sampled sequence $y_i (i = j, \ldots, k)$ about the mean value μ;

$S_j^k(\mu_1)$ = detection statistic, which is the log likelihood ratio of the joint frequency function for the independent variables y_i during time j to k about μ_1 and μ_0.

The behavior of the detection statistic in the absence or presence of a step change in electrical power can be easily understood. Suppose that the function is used to test whether a chiller of known electrical power has turned on. Before the event to be tested, power data, collected at a measurement point that includes electrical service to the chiller, are distributed about the pre-event mean power level. If the chiller

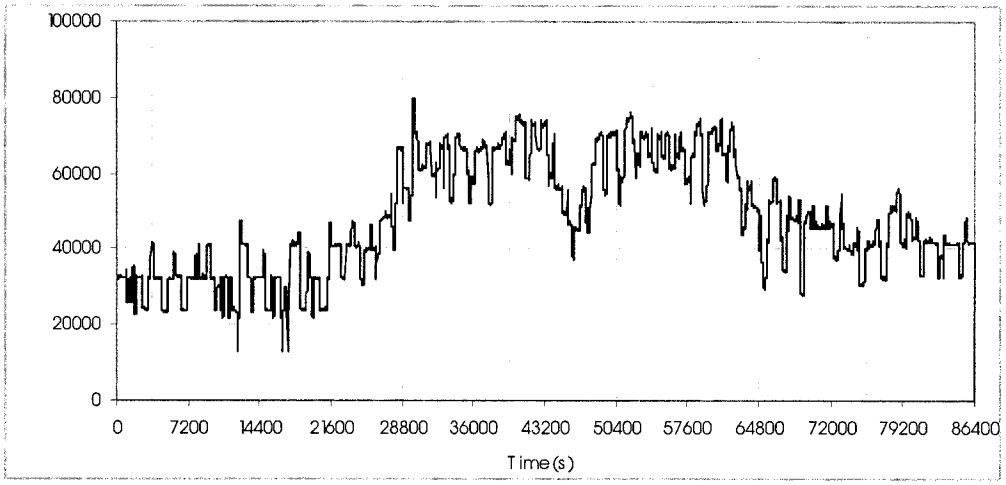


Figure 1 *Whole-building electrical power, sampled at 24 Hz. Data were collected at the electrical-service entry to the test building used in ASHRAE RP-1020.*

does not turn on, the power measurements will not change. The probability density function (PDF) in the numerator of Equation 1, which is centered on a post-event mean that includes the chiller power, will be very small, while the PDF in the denominator will be much larger, because data points will still be clustered about the pre-event mean. In other words, there is a very small probability that the measured power level would be associated with the operating chiller. The probability ratio is therefore very small and the log of the ratio is large and negative. As the magnitude of the hypothesized change in power decreases, the ratio approaches a value of 1.0 in the absence of a power change and the natural log approaches a value of 0. By contrast, if the chiller turns on, the PDF in the numerator will be large, because data points will be centered about the hypothesized mean, and the PDF in the denominator will be small. The ratio, therefore, is large.

There are two independent variables, the change time and the mean value after the change, which leads to a double maximization of the detection statistic S_j^k:

$$g_k = \max_{1 \le j \le k} \ln \hat{\Lambda}_j^k = \max_{1 \le j \le k} \sup_{\mu_1} S_j^k(\mu_1) \quad , \tag{2}$$

where

$\hat{\Lambda}_j^k$ = estimate of the upper bound of the ratio of the joint frequency function about the post-event mean value μ_1 for a given pre-event mean μ_0, within the window [*j*, *k*].

sup = supremum, i.e., the least upper bound of S_j^k over [*j*, *k*] about the mean value μ_1 with reference to the known mean μ_0 before the change.

In other words, the event identified by the maximum probability ratio is found by searching for the time *j* and the corresponding average μ_1 in each subwindow *j-k* in the current detection window *1-k*.

In some cases, the minimum magnitude of the change in power for a given component is known in advance. The minimum value of the change of parameter μ, designated as V_m, can be used in the search of μ_1, i.e., $|\mu_1 - \mu_0| \ge V_m$. Equation 2 can then be rewritten as:

$$g_k = \max_{1 \le j \le k} \sup_{|\mu_1 - \mu_0| \ge V_m \ge 0} S_j^k(\mu_1) \quad . \tag{3}$$

Because noise in electrical power of equipment follows the normal distribution, according to the central limit theorem (Rice 1988), noise in the total power data of a system consisting of components independent in power consumption can also be described by the normal distribution. Therefore, the deviation of the sampled total power data y_i from the calculated mean μ can be represented by the normal distribution, $N \sim (0, \sigma^2)$. Here, σ stands for the standard deviation of the total power data of the monitored system.

For an independent Gaussian sequence, the probability density function is

$$P_\mu = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(y_i - \mu)^2}{2\sigma^2}} \quad . \tag{4}$$

The detection statistic can then be derived as

$$S_j^k = \frac{\mu_1 - \mu_0}{\sigma^2} \sum_{i=j}^{k} \left(y_i - \frac{\mu_1 + \mu_0}{2} \right) . \tag{5}$$

Let $V = \mu_1 - \mu_0$, the change in mean power signal after an on-off event. Then

$$g_k = \max_{l \le j \le k} \sup_{|V| \ge V_m > 0} \sum_{i=j}^{k} \left[\frac{V^*(y_1 - \mu_0)}{\sigma^2} - \frac{V^2}{2^*\sigma^2} \right] \quad (6)$$

and

$$\hat{V}_j = \left(\frac{1}{k-j+1} \sum_{i=j}^{k} |y_i - \mu_0| - V_m \right)^+ + V_m \quad (7)$$

where $\hat{V}_j$ is the value of V at which g_k reaches its maximum,

$$\hat{V}_j = \begin{cases} V_m & \text{if } \frac{1}{k-j+1} \sum_{i=j}^{k} |y_i - \mu_0| < V_m \\ \frac{1}{k-j+1} \sum_{i=j}^{k} |y_i - \mu_0| & \text{if } \frac{1}{k-j+1} \sum_{i=j}^{k} |y_i - \mu_0| \ge V_m \end{cases}.$$

Then g_k can be rewritten as

$$g_k = \max_{l \le j \le k} \sum_{i=j}^{k} \left[\frac{\hat{V}_j^*(y_1 - \mu_0)}{\sigma^2} - \frac{\hat{V}_j^{\,2}}{2^*\sigma^2} \right]. \quad (8)$$

If $V_m = 0$, meaning that a change of mean of any magnitude is of interest or no information about the minimum expected change is available in advance, then

$$g_k = \frac{1}{2^*\sigma^2} \max_{l \le j \le k} \frac{1}{k-j+1} \left[\sum_{i=j}^{k} (y_i - \mu_0) \right]^2. \quad (9)$$

The magnitude of g_k increases with the change in power and the abruptness of the change. A value above a threshold indicates an on-off event of potential interest.

Training the Parameters of the GLR Algorithm

The GLR detector requires that four parameters be trained for a given application:

1. the length of the pre-event averaging window;
2. the length of the detection window;
3. the threshold for the detection statistic; and
4. the standard deviation (or variance) of the power data.

Guidelines follow for selecting the parameters in one or more of three ways: *a priori* experience, tuning the parameters on the basis of one-time measurements, or automatically adjusting the parameters. The first two guidelines concern the size of sliding windows, as expressed in a number of samples. These guidelines have been shown to be reasonable for a range of sampling rates. Guidelines for the sampling rate itself will be presented later in this paper.

The Length of the Pre-event Averaging Window. Because multiple power changes occur in sequence in HVAC systems, the mean before the change must be continuously updated with each new data point. The length of the data window used to estimate the pre-event average power has a profound impact on the detection statistic. The GLR algorithm with a short window will yield a large detection statistic when the incoming data include spikes, which will be a source of false alarms if the detection statistic exceeds its threshold value. Further, the GLR algorithm with a short window may also miss events that occur gradually over several data points. On the other hand, a GLR detector with a long window will not find multiple abrupt changes that are close to each other in time.

On the basis of trial-and-error tuning, the appropriate length of the pre-event averaging window was 10 data points for a set of one-minute-averaged data that recorded electrical power for four air handlers, each consisting of a supply and a return fan. At the ASHRAE test building, this window was set to be six data points, sampled at one-second intervals. An estimate of a suitable window length can be made on the basis of the characteristics of the HVAC components, how often they are switched on or off, the status of the electrical facility, and the typical power profile of the system. Observations made in this research indicate that the upper limit for the length of the pre-event averaging window should be no longer than the interval between two major consecutive events. This interval can be estimated from a basic knowledge of the monitored HVAC system. Major individual components usually have minimum on and off times, to protect against deterioration due to frequent on/off switching. Moreover, systems of components are or can be controlled to operate in a sequence with specified time intervals, as in turning on or off air-handler fans. Note that this requirement is based in the time domain and is converted to a number of samples on the basis of the sampling rate.

As a lower limit, the pre-window should never be shorter than significant electrical noise spikes and ideally should be longer than the duration of the start-up period of each single component in the system. However, this condition cannot always be met for a real system, because the start-up of a VSD fan can be as long as 15 minutes while the interval between the on/off transitions of two components can be shorter than this. In such cases, other approaches are needed, such as the multi-rate sampling technique discussed later. Practical experience has shown that the window should contain at least four data points.

Without violation of the above basic rules, the pre-window should be kept short. If there is a range of possible intervals between the lower and upper bound, a value at or near the lower bound should be selected. For a given system, the window length for detection seems to be consistent across different seasons and equipment operating conditions, as demonstrated by extensive tests to detect HVAC equipment on-off events in the ASHRAE 1020-RP test building (Shaw et al. 2002; Norford et al. 2000, 2002).

If the multi-sampling rate detection algorithm (to be described later) is used, the number of data points in the window should be determined from the above time-domain guidelines and the fastest sampling rate. The same number of

data points are used at all sampling rates in the multi-rate algorithm.

The Length of the Detection Window. The above basic limits for the pre-event averaging window also apply for the post-event detection window—i.e., it should not be longer than the interval between two consecutive events, never shorter than a disturbance, and ideally not shorter than the duration of a start-up transient process. On the other hand, unlike the pre-event averaging window, which is used to achieve a stable mean as the reference for coming events, the post-event detection window is intended to be sensitive to events yet robust to disturbances. A shorter post-event detection window is more sensitive to changes than a longer one. Moreover, to reduce the computing time required to search for a change in the post-event detection window, its length should be as short as possible. The appropriate length of the post-event detection window was found to be 25~50% of the pre-change average window in order to get a relatively stable yet sensitive average for detection of on-off events. For the fan data sampled at one-minute intervals, the detection window was set to five data points, 50% of the pre-event averaging window. For the ASHRAE test building, the detection window was three data points, also 50% of the pre-event averaging window.

The Threshold for the Detection Statistic. Literature on the GLR method describes the detection threshold, a dimensionless quantity, as a trained parameter. The magnitude of an appropriate detection threshold scales with signal noise, the minimum signal change of interest (which comes from a knowledge of rated equipment-power levels), and the abruptness of potential changes in the system. For detection of on/off events from the total electrical power of eight fans in a campus building, the threshold was set to 200 on the basis of on/off tests. The same threshold was used at the ASHRAE test site. For general detection applications, this threshold might be set adaptively during tests designed to determine detection parameters appropriate for a given HVAC system. The threshold for the detection statistic can be initiated as an arbitrary small number, for example a magnitude of one, and then increased until all events of interest are identified with a minimum occurrence of false or missed alarms.

The Standard Deviation of the Power Data. The standard deviation is an important measure of data quality, which for HVAC-system electrical-power data may vary rapidly over time due to noise and changes in power of equipment of interest. In the GLR algorithm, the magnitude of the detection statistic, as shown in Equation 8 for the case where power changes of any magnitude are of interest, is inversely proportional to the standard deviation. Therefore, calculation of the standard deviation becomes a key issue for successful change detection. The simplest method to determine the standard deviation is to measure it on a one-time basis during a training period. However, GLR output based on a fixed standard deviation was rarely fully satisfactory, which prompted one of the improvements to be discussed in the next section.

The standard deviation is valuable information of itself, in addition to its impact on the GLR detection statistic. It will increase noticeably when a fan, pump, or chiller under closed-loop control is unstable due to poorly selected controller gains. Norford and Leeb (1996) showed that a poorly tuned chiller controller could be seen in the electrical signal measured at the HVAC service entry. Detection of the oscillations can be automated via calculation of the standard deviation in the data window.

IMPROVEMENTS TO THE DETECTION ALGORITHM

Resetting the Detector after an On/Off Event Has Been Detected. Consider a single on/off event, detected by the GLR algorithm as it works its way through a continuous stream of data, sliding the pre-event mean window over one data point at a time and then calculating the detection statistic for each time in the detection window. When the detection statistic exceeds the threshold, it is desirable that the alarm be immediately silenced. That is, a single event should produce a single needle spike in the detection statistic, because an alarm of appreciable width can mask subsequent events. However, close examination of the algorithm reveals that it can continue to produce alarms as the windows slide through the data. Only when the windows move entirely past the on/off event will the alarm necessarily cease. This phenomenon clearly depends on the length of the windows, the abruptness of the event, signal noise, and the value assigned to the threshold. While it is possible to adjust window width to account for this effect, a better strategy is to purge the pre-event mean window when an alarm first occurs and refill it with new data. At the time of purge, the detection statistic will immediately drop down below the threshold. Figure 2 shows a representative pre-event averaging window, detection window, and reset window, used to reload the pre-event averaging window. The length of the reset window is, of course, the same as the pre-event window.

Updated Standard Deviation. The standard deviation can be continuously calculated from data in the pre-event averaging window. While such an on-line calculation is more accurate than a single calculation during the GLR setup period, it was found necessary to limit the range of allowable values because such extremes unduly influence the magnitude of the detection statistic. For example, a very steady, noise-free electrical signal will have a low standard deviation and the detection statistic will be very large, even for events associated with very small power changes that are of no interest. A very noisy electrical-power signal will cause the detection statistic to be very small, potentially masking events of interest.

Limits on the standard deviation were assigned as a fraction of the total power. Tests showed that the standard deviation tends to increase as more equipment is in operation and the total power increases. Data from a training period were used to determine the ratio of the measured standard deviation and the measured total power. During subsequent on-line FDD tests, the standard deviation was estimated as a product of this

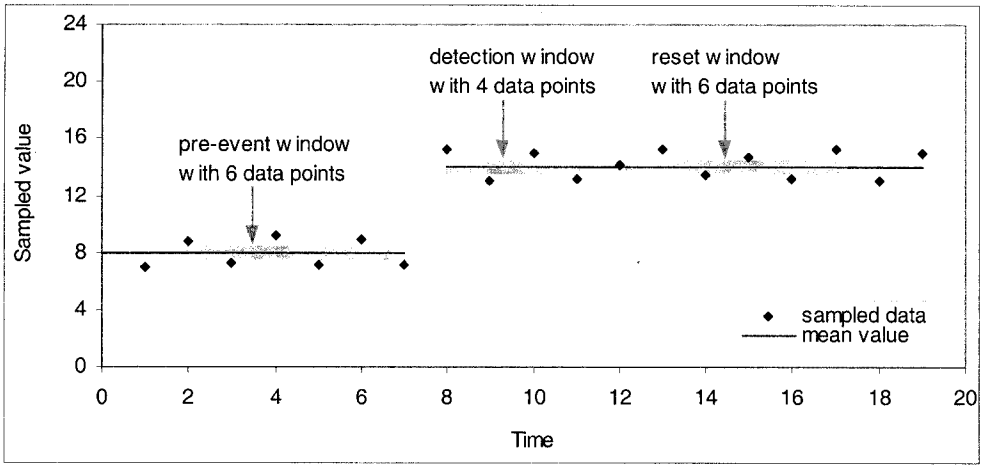


Figure 2 *Data windows used with the GLR detection algorithm.*

ratio and the measured total power. This method has given the most reliable estimate of the standard deviation because it eliminates the effect of the extreme values of the standard deviation while incorporating an updated estimate of the standard deviation in the calculation of the GLR detection statistic. At the ASHRAE test site, reasonable upper and lower limits for the standard deviation as fractions of the current power data were taken to be 10% and 1% respectively.

Non-Zero Minimum Expected Change. The GLR equation is easy to implement when the minimum expected change in power is zero. However, the zero-power minimum may cause more false alarms than a minimum value assigned on the basis of knowledge of equipment size. This can be readily seen from Equation 8: g_k increases with decreasing $\hat{V}_j$. In practice, it is often reasonable to find and set some minimum expected change based on the knowledge of the system and its components. For example, if the NILM were used to analyze fan performance after a variable-speed-drive retrofit, the minimum power level of interest would be set by the fans and would exclude smaller pieces of equipment, such as chilled-water pumps, as well as unknown disturbances. With the properly determined minimum expected change V_m, the detection statistic is insensitive to such disturbances and their accumulation within one window length, thus making the detection more reliable. A minimum expected change of 200 W was selected for the ASHRAE test building.

Median Filter. The performance of the GLR is adversely affected by signal noise. One way to reduce the impact of noise is to pre-process the data with a median filter (Karl et al. 1992), which simply picks the median value of the sequence to represent the current value. The length of the median filter's window should not be shorter than the duration of an electrical spike, observed in this study to be generally less than five seconds, and not longer than the interval between two consecutive events, selected as 20 seconds. The median filter is designed for the base sampling rate or interval, in the event multi-rate sampling is used. For example, if the base sampling interval is one second, then the number of data points in the filtering window should be between 5 and 20. The filtered data are then used for detection with different sampling rates, if necessary. At the ASHRAE test site, ten data points were included in the median filter window.

One possible problem with the median filter is that it masks rapid power oscillations. At the ASHRAE test building, this potential masking was avoided by using different data sets, one for change detection with the power data processed by a median filter and the other for oscillation detection without the median filter.

Combined Detection of On/Off Power Changes and Power Oscillations. Oscillation of power caused by unstable control in HVAC systems may degrade equipment and in some cases increase energy consumption, and is a fault that should be detected. One of the major characteristics of oscillations is the deviation of the data from their mean value in a window of appropriate length. This fault was identified by comparing the standard deviation of the data against a threshold that was dynamically adjusted as a fraction of the current power data.

The key points in designing a GLR detector with oscillation detection capability are proper thresholds for the detection statistic and for the standard deviation. Detecting step changes in power requires an upper threshold for the detection statistic, HGLR (high value for GLR), and a lower threshold for the standard deviation, LSTD (lower value for standard deviation). To detect an on-off event, the GLR detection statistic must exceed HGLR and the standard deviation must be no larger than LSTD. The oscillation detector relies on an upper threshold, HSTD, for the standard deviation and a lower threshold, LGLR, for power changes. The standard deviation must exceed HSTD and the GLR detection statistic must be no larger than LGLR. Thresholds established in this manner eliminate the fuzzy intermediate region where, for example, fluctuations in power may cause false alarms in the change detector, and permit a clearer distinction of step changes and

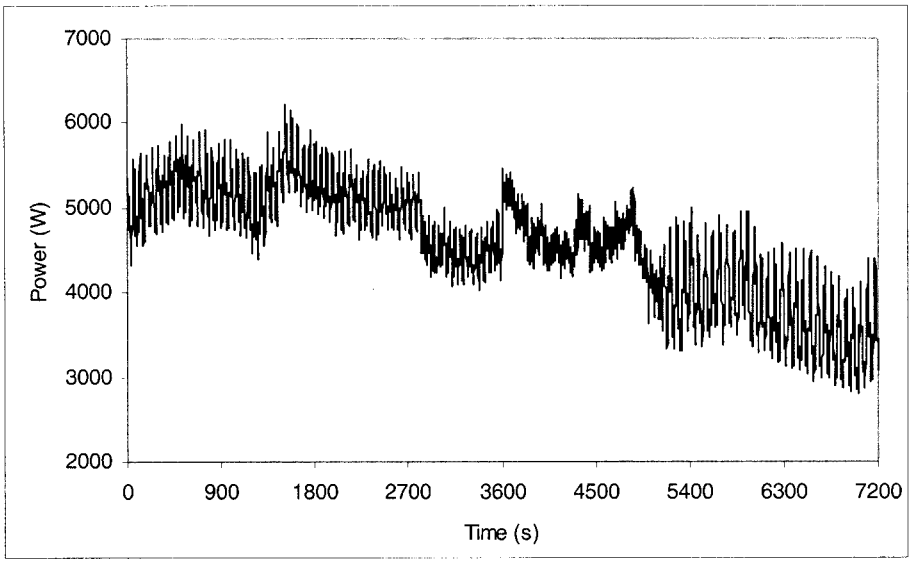


Figure 3 *Total electrical power of the HVAC system in a test building, showing oscillations indicative of an unstable supply-duct static-pressure controller as well as an on-off switching event.*

oscillations. The LGLR and the LSTD were set to be less than 10% of the HGLR and 20% of the HSTD, respectively, and were trained with measured data to minimize false alarms. Figure 3 shows data with both on-off events and an unstable controller, using power data recorded by a NILM installed at the motor-control center in the ASHRAE test building. During a period when the supply-fan static-pressure controller gain was set to produce power oscillations, a chilled-water pump was turned on at about 3600 seconds. With HGLR= 2500, LGLR= 250, HSTD= 0.02, and LSTD= 0.004, both events were alarmed successfully.

OPERATION OF THE GLR DETECTOR WITH MULTIPLE SAMPLING RATES

In spite of the improvements just described, applying the GLR detector to data sampled at a single rate inevitably involves a conflict between sensitivity to rapidly occurring events and susceptibility to false alarms generated by electrical noise. It is therefore worthwhile to explore the benefits of applying the GLR to a data set sampled at different rates. This will be done in three steps. First, the rules established above for determining the sampling rate will be applied and will be shown to have unavoidable limitations in detecting rapidly sequenced events. Second, the same rules will be shown to be unsatisfactory in detecting chiller on/off cycling in a noisy electrical signal. Third, the GLR detector will be applied at different sampling rates and the results combined in a way that overcomes some of these limitations.

Order of Magnitude Analysis of Sampling Rate Selection

Selecting a single sampling rate requires considerable care. An appropriate starting point is to bound the sampling interval: the upper bound should be less than the shortest time interval between two consecutive events of interest and the lower bound should be greater than the time required for the fastest turn-on event of interest. The need for the upper bound is clear, but it may lead to sampling rates that leave the GLR sensitive to noise, as will be discussed shortly. The lower bound is needed because shorter sampling intervals will make a step-change take on ramp-like properties and be harder for the GLR algorithm to detect.

The interval between consecutive events can be very short, as illustrated with a single day of whole-building data, taken at the test building used for ASHRAE 1020-RP and previously shown in Figure 1. The data were collected with a NILM and were sampled at 24 Hz in the first stage of the analysis. During a single 30-second period (12300-12330 seconds) three devices were turned on sequentially. (The identity of these devices is not important for this analysis and was not determined. The whole-building NILM was used primarily to detect the reciprocating chiller, and other events were not classified.) Reducing the sampling interval from 60 to 10, 1, and finally 0.125 seconds improved the resolution. Figure 4 shows this progression.

The three turn-on events were taken to be a single step change at a sampling interval of 60 seconds. With the sampling interval reduced to 10 seconds, the events were shown as single data points of different magnitude, discernible to the eye but not to the detection algorithm. When the sampling interval was reduced to 1 second (not shown in

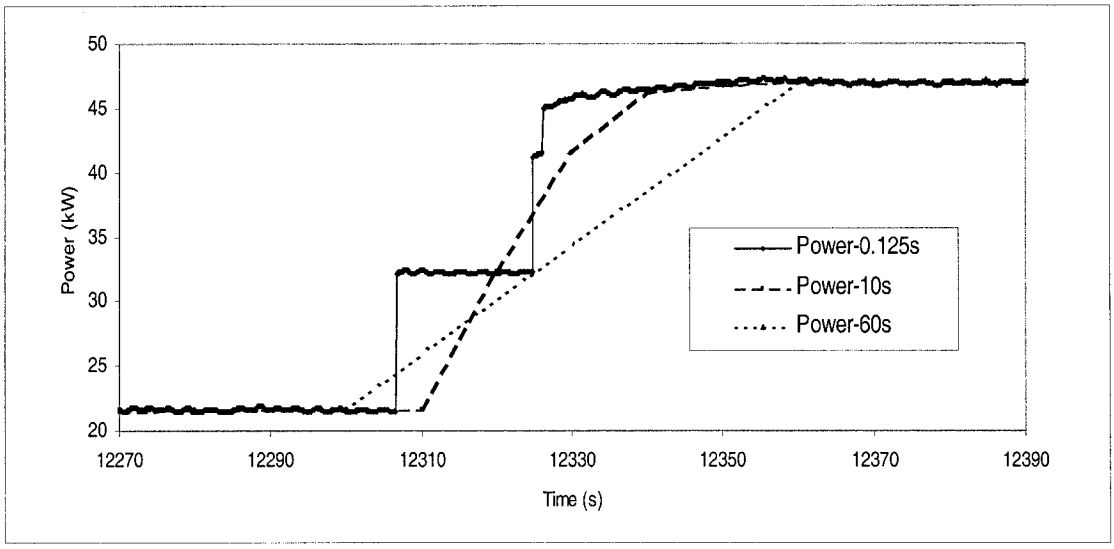


Figure 4 *Power data sampled at three different intervals during a period of 120 seconds when three events occurred, demonstrating the significantly different data patterns fed to the detector due to the varying sampling intervals.*

Figure 4 because it nearly overlapped data taken at 0.125 second intervals), the first and second events could be clearly recognized by eye and by the detection algorithm as well. The last event was discernible to the eye but not to the detection program, which needed an appropriate stream of data samples to build up the GLR detection statistic. The last event remained ambiguous to the program until the sampling rate was increased to 8 Hz, i.e., an interval of 0.125 seconds.

At the ASHRAE test building, the duration of the fastest turn-on event was 0.125 second and the shortest time interval between events was 1.25 seconds. As just shown, a sampling interval of 0.125 second was needed to produce enough data points to distinguish start-up transitions separated by 1.25 seconds. But subsequent tests to detect cycling of a reciprocating chiller showed that the false alarm rate increased drastically when the sampling interval dropped below one second. Because the three unknown start-up signals were not of interest and reliable detection of the chiller was central to the research, the minimum sampling interval was set to one second. Events spaced more closely than the sampling rate were unavoidably missed or misinterpreted. Such events seem to happen by coincidence and are not often seen in an HVAC system. In tests to date, there was generally an interval of not less than 10 seconds between on- and off-transitions for such coupled equipment as supply and return fans associated with the same air handler. The slower sampling rate not only reduced false alarms but also reduced the execution time of the GLR detector. This was important when the GLR detector was operated with multiple sampling rates as detailed later. The running time of the detector, as implemented on a personal computer with a clock speed of 233 MHz, was reduced from 90 minutes at an 8 Hz sampling rate to five minutes at a 1 Hz sampling rate for the evaluation of a single day's test, making the detector more desirable for on-line detection with comparable computers.

Detection of On/Off Changes with a Single Sampling Rate

Different sampling rates were used to examine power data for six days from the ASHRAE test building, with the goal of detecting reciprocating-chiller cycling. Figure 5 illustrates the first seven hours of the total power data on a single day. Figure 6 shows chiller power as measured with a submeter as well as GLR detection output for the chiller for a single sampling rate of 1 Hz. Note from Figure 5 that it is not easy to visually distinguish the chiller cycles from the whole-building power signal. The first chiller cycle, aligned with data from the submeter, is readily discerned. Most of the following events are obscured by other events of equal or greater magnitude.

The GLR detector identified all 13 start-up events from the 1 Hz samples but missed four of the shut-down events, which were masked by concurrent events. Other sampling intervals, including 2, 3, 4, 5, 6, 7, 8, 10, 12, 15, 20, 30, 40, 50, and 60 seconds, were also tested. However, all those sampling intervals produced results similar to those shown in Figure 6. With any one single sampling interval, the detector was not able to find all the on/off switches correctly. However, it was possible to visually identify all of the events by matching the on's and off's among outputs with different sampling rates. In this case, on/off matching among the outputs with the sampling rates of 1, 2, and 5 seconds identified all the on/off switches without any false alarms or missing events. Such other combinations as 1, 2, and 10 seconds were also successfully used for the matching. Similar detection patterns with single sampling rates were found with the remaining five

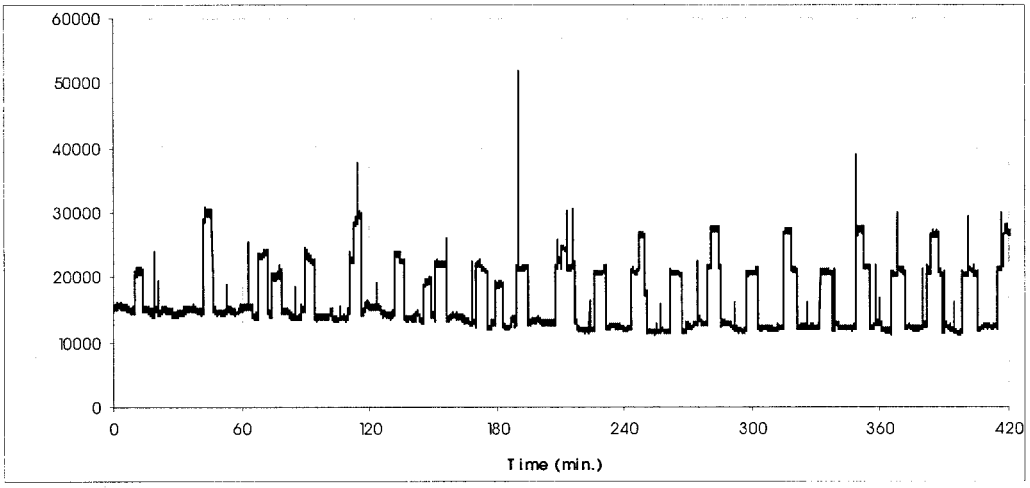


Figure 5 *Whole-building power data from the ASHRAE 1020-RP test site, sampled at 24 Hz and plotted at 10-second intervals. The time period is about seven hours.*

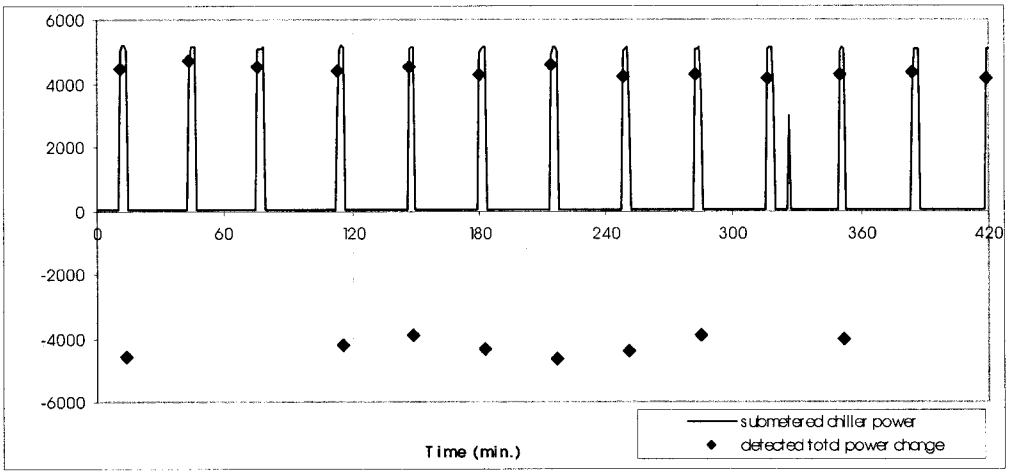


Figure 6 *Submetered chiller electrical power and the on/off cycles detected by the GLR algorithm as changes in the building's total power data sampled at 1 Hz over the same time period as in Figure 5.*

days' data. This prompted an effort to automate the process of matching events detected at different sampling rates.

Detection of On/Off Changes of Specific Equipment by Automatic Matching Among Multiple Sampling Intervals

The basic GLR detector was extended to analyze data at multiple sampling rates. Each sampled series with a specified integer-sampling interval between the lower and upper limits was supplied to the GLR detector. For the ASHRAE test building, an acceptable maximum value for the sampling rate was about an order of magnitude longer than the minimum value. Specifically, the minimum sampling interval was one second and the maximum was set to be 30 seconds.

Once an event was found by the detector, the power change was calculated as the difference between the mean in the post-event window and that in the pre-event window. The mean in the pre-event window is the average of data in the whole window, while that in the post-event window is calculated with the data from the point where the event is found to the end point of the window. The detected power changes for each sampling interval were sorted by the time of the changes, assigned to the equipment with that magnitude, and matched with the output from other sampling intervals. Matching of events involved the time of event and the sign (positive for on-transitions and negative for off-transitions) and the magnitude of the changes.

Upper and lower limits need to be trained for the magnitude of change or set adaptively (Hart 1992). Such limits can be trained by applying the detector to data collected during one day's operation. For the detection shown in Figure 6, for example, data during the first seven hours of a normal operation day will suffice for training. At some time point, if a positive change within the given limits is found, then the next negative change within the magnitude limits seen by any of the employed sampling rates will be assigned to the positive change and an on/off cycle is recorded.

Data filtering criteria may be applied to specific equipment to reduce the false alarm rate. This is helpful when the power of a component of interest is similar to that of other equipment, as is the case at the test building (Figure 5). For the chiller, for example, these criteria include the minimum off-time between cycles, which is typically set within the chiller controls to prevent unnecessary equipment cycling, and the minimum expected on-time. These limits were incorporated in the chiller-detection algorithm. Such filtering criteria are based on normal, not faulty, chiller operation. The above "minimum" criteria can be used for the detection and diagnosis of such faults as short cycles or prolonged operation time. This can be done with two different output files, one using the "minima" for filtering and the other using them for fault detection and diagnosis. If an excessive number of equipment cycles is detected in an unfiltered data stream, then there is either a fault or there are other components of comparable magnitude. If the power magnitudes of different equipment are distinct from each other, then these "minima" will not be needed for filtering and can be limited to detection of cycling faults.

The multiple-sampling-rate GLR approach has proved to be very useful in coping with electrical-power complexities not found in residential buildings. First, start-up and shut-down events vary in their apparent abruptness, as a function of the equipment (soft-start motors, for example) and as influenced by changes in other loads. In the ASHRAE test building, VAV air-handler fans have a very slow start-up signature because they are controlled by variable-speed drives. With the multi-rate sampler, these start-up transients can be detected, as was shown for the VAV fans in the test building. Detecting the fans required that the maximum sample interval be increased from 30 seconds to 10 minutes. At the other end of the sampling-rate spectrum, abrupt shut-down transients are difficult to detect when masked by gradual changes in power drawn by other equipment. In these cases, the fast samplers included in the multi-rate algorithm work best. Second, as shown in Figures 7 and 8, power oscillations characteristic of a poorly tuned controller were more reliably detected via analysis of data sets taken at multiple rates. This approach tends to mitigate the problem of picking a sampling rate appropriate for detection of oscillations at an unknown frequency.

SUMMARY OF TRAINING GUIDELINES

Guidelines for training the GLR detector, presented in detail above, are summarized as follows:

1. Record electrical power for the circuits monitored by the NILM for one day under typical operating conditions. The sampling rate should be between 1 and 10 Hz for common HVAC systems.
2. Locate the events from the abrupt changes in the total power data and estimate the fastest and the slowest events.
3. Determine the base sampling rate for detection. The base sampling rate will be used as the fastest sampling rate if multi-rate sampling is employed. Therefore, power data sampled at this rate should be used to discern by eye each event of interest.
4. Determine the window lengths. The length of the detection window should contain at least two data points. It should not be longer than the interval between two consecutive events and never shorter than a disturbance. The length of the pre-event window is 2~4 times longer than the detection window. The length of the post-event window is the same as that of the pre-event window.
5. Calculate and estimate the standard deviation of the power data as a fraction of the current total power data for the detection, f. The lower/upper limits of the standard deviation can be set at 1~2 magnitudes lower/higher than the estimated value f.
6. Estimate the threshold for the detection statistic. A reasonable base value for the threshold is $1/f^2$. The threshold can then be trained by adjusting this value until all events of interest can be seen by the detector, with a minimum number of false alarms.

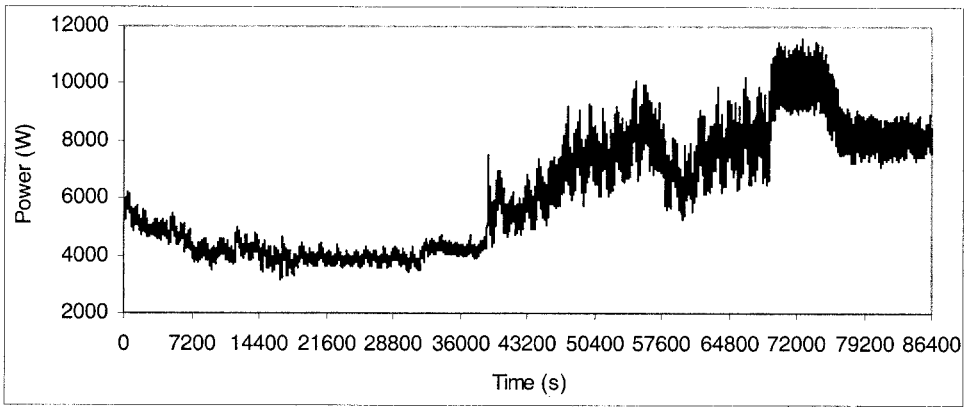


Figure 7 *Motor-control-center electrical-power data from the ASHRAE test site, taken over a 24-hour period beginning at 8 p.m. The controller gain for the supply fan in one of three VAV air handlers in the building (and monitored by the NILM on the motor-control center) was increased at about 7:25 a.m. to the point where the fan control was unstable.*

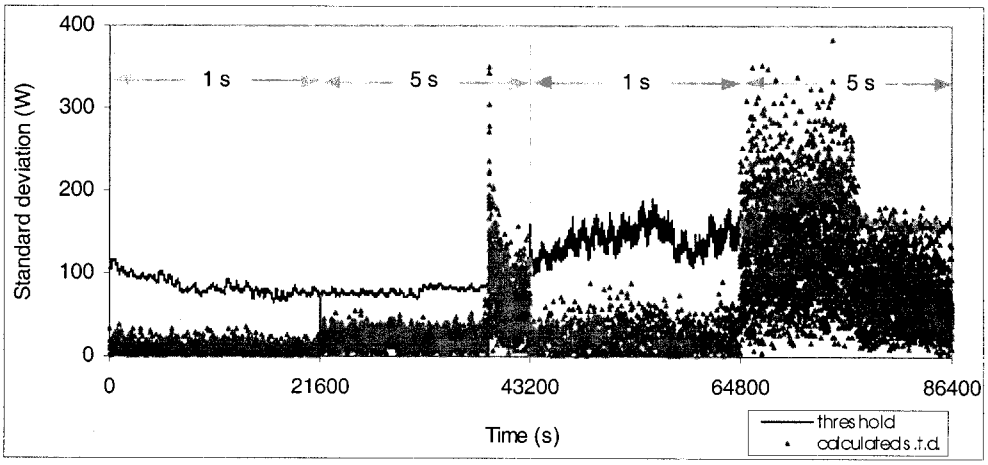


Figure 8 *Detection of the unstable supply-fan controller via calculation of the standard deviation with data sampled from the motor control center at two different intervals, 1 and 5 seconds, combined and shown in different time sections of 24-hour operation.*

It should be noted that exact values are not expected for the above parameters due to the statistical properties of the detection method. Slightly different combinations of these parameters may yield equally acceptable detection output.

The detector developed in this research has been successfully applied to another test site. With the above basic rules and guidelines, the training process for the parameters became much easier. Because the rules for the window lengths are for common HVAC systems, the same window lengths were used and the training was confined to determination of the thresholds for the detection statistic and the standard deviation. On/off events at this site were more abrupt than at the ASHRAE test building, in part because there were no variable-speed drives, and the detector produced satisfactory results while operating at a single sampling rate.

COMPARISON OF ON/OFF CHANGE DETECTION WITH CENTRALIZED AND SUBMETERED POWER DATA

A well-trained GLR detector will detect a high percentage of on/off events of interest, with a minimum number of false alarms. However, such performance does not necessarily mean that power changes can be quantified with sufficient precision to detect potentially faulty conditions. Table 1 and Figures 9-11 compare the GLR with single and multi-rate sampling against data from electrical submeters at the ASHRAE test building. During this five-hour test period, which started at 19:44 and included normal equipment operation and the end-of-day shut-down sequence, there were 17 on-off switching events for two hot-water pumps, two supply fans, and two return fans, among a total of six fans and 10 pumps served by the monitored circuits.

TABLE 1
ON-OFF POWER CHANGE DETECTION FOR A MOTOR CONTROL CENTER SERVING FANS AND PUMPS IN A TEST BUILDING, AS COMPARED WITH SUBMETERED DATA

Time	ON/OFF Equipment	Submeter (W)	NILM - GLR			
			Single Interval		Multiple Interval	
			Power (W)	Error (%)	Power (W)	Error (%)
20:10	Loop-B hot water pump—OFF	–405	–502	19.3	–400.6	1.1
20:19	Loop-B hot water pump—ON	264	249	6.0	284.1	7.1
20:21	Loop-A hot water pump—OFF	–275	Not found		–279.0	1.4
20:25	Loop-A hot water pump—ON	252	313	19.5	226.5	11.2
20:34	Loop-B hot water pump—OFF	–244	–98	149.0	–476.1	48.8
20:44	Loop-B hot water pump—ON	232	497	53.3	230.4	0.7
21:02	Loop-B hot water pump—OFF	–309	–346	10.7	-281.6	9.7
21:14	Loop-B hot water pump—ON	267	171	56.1	252.1	5.9
21:29	Loop-B hot water pump—OFF	–202	–11	1736.4	–171.5	17.8
21:47	Loop-B hot water pump—ON	237	289	18.0	257.2	7.9
21:55	AHU-A supply fan—OFF	–1400	–1396	0.3	–1410.9	0.8
21:56	AHU-A return fan—OFF	–82	Not found		Not found	
22:00	AHU-B supply fan—OFF AHU-B return fan—OFF	–430	–274	56.9	–485.1	11.4
22:06	Loop-B hot water pump—OFF	–300	–242	24.0	–243.2	23.4
22:26	Loop-A hot water pump—OFF	–264	–283	6.7	–244.4	8.0
23:40	Loop-A hot water pump—ON	342	373	8.3	220.1	55.3
00:33	Loop-B hot water pump—ON	295	494	40.3	450.5	34.5

With Single Sampling Interval

Of the 17 events, 15 were found with different levels of error and the remaining 2 were missed. Table 1 and Figure 10 show that the quality of the detection depended on the magnitude of the change relative to the total power as well as the current data trend. For example, when water pump B turned off at 20:34 and the change in power of –244 W was less than 5% of the total power, the error was about 150%, relative to the submetered change in power. However, when the supply fan for AHU-A turned off at 21:55 and the power change of -1400 W was about 25% of the total power, the error was less than 0.3%, again relative to the submetered power. When the change magnitude was very small, the detector was not able to find the event at all, including the missed event at 21:56 when the return fan for AHU-A was turned off with a power change of only –82 W.

The shorter the time interval between two changes, the more difficult it was to find the changes, especially small ones. This is simply because as a steady-state detector, the GLR needs some time for the effect of the former event to die out in order to find the subsequent change. The smaller the magnitude of the change, the longer the time interval needed to detect it. This is demonstrated by the output at 20:21 for the turnoff of pump A. With a magnitude of -275 W, the event was

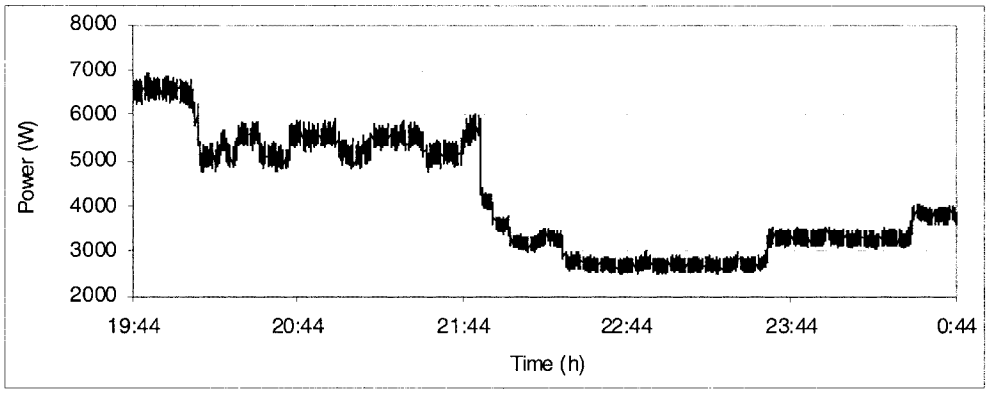


Figure 9 *Electrical power data collected at the motor-control center in the ASHRAE 1020-RP test building, taken over a five-hour interval beginning at 7:40 p.m. that includes the normal evening shutdown period. Seventeen on-off transitions are included in the data.*

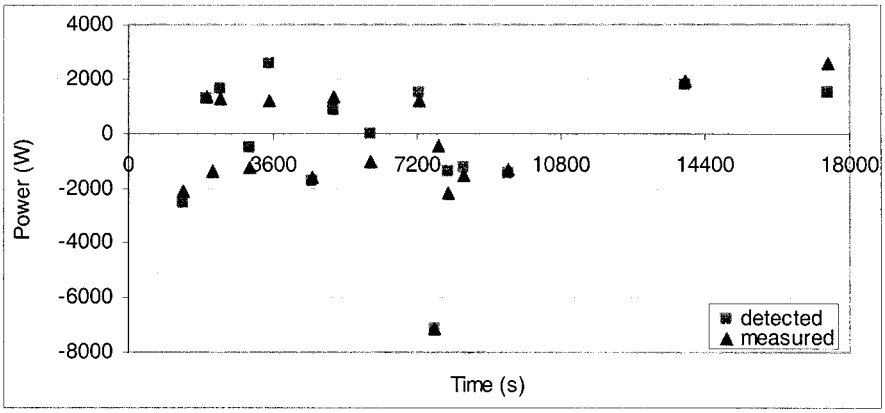


Figure 10 *Comparison of centralized and submetered power monitoring in the ASHRAE 1020-RP test building, with analysis of data taken at a single sampling rate of 0.1 Hz.*

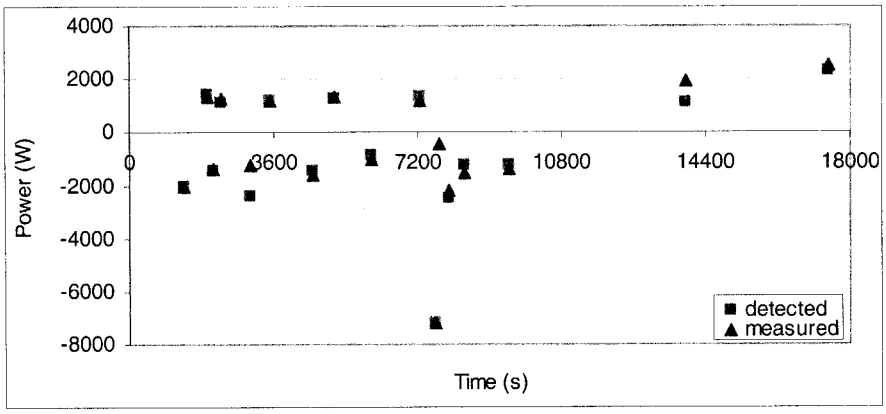


Figure 11 *Comparison of centralized and submetered power monitoring in the ASHRAE test building, with analysis of data taken at multiple sampling rates. Power monitoring is improved relative to analysis of data at a single sampling rate of 0.1 Hz.*

still missed because it was masked by gradual changes in electrical power due to the previous off-transition and to noise, which can be seen in Figure 9.

With Multiple Sampling Intervals

The multiple-sampling-rate approach has been shown to improve both the detectability of events and the resolution of the associated change in electrical power. This can be demonstrated by running the detector with multiple sampling intervals through the same five-hour data set used to evaluate the single-sampling-rate algorithm and plotted in Figure 9. As shown in Figure 11, the multiple-sampling-rate detector identified 16 of 17 on/off events, one of which was undetectable with the single-sampling-rate approach. Moreover, for almost all events the detector estimates a change in power that is closer to the submeter measurements than was the case for the single-sampling-rate algorithm, as listed in Table 1.

Under a few circumstances, the detection error for the multi-rate sampler was larger than that for the single sampling rate. With the lowered threshold of the detection statistic for multiple rates (the lowest value among the different sampling rates), the data pattern for reset of the detection window is different from any single sampling rate with a higher threshold, and this causes some variation of the change magnitude. But the occurrence of such problems is rare because the multi-rate detection is virtually a vote among different sampling intervals.

The major issue to be addressed in the near future for the multi-rate detection is to develop an appropriate algorithm to automatically calculate or select the "optimal" change value among different sampling intervals, i.e., the value that is closest to the real change. This might be realized by averaging, voting, or sorting among given sampling intervals on the basis of observations with more testing data.

DESCRIPTION OF NILM HARDWARE

A block diagram of the non-intrusive load monitor installation at the test facility used in ASHRAE RP-1020 appears in Figure 12. In addition to the conventional NILM-style connections at the main panel (for remote1) and the motor control center (for remote2), selected loads associated with air-handling unit B and the chiller were individually measured. With the connections shown in Figure 12, remote1 measured spectral envelopes (i.e., power envelopes for the fundamental and higher harmonic signals) for the entire building and remote2 measured spectral envelopes for the motor control center. Remote1 and remote2 shared the task of collecting data from individual loads.

The hardware platform for the prototype event detector consisted of a personal computer working in tandem with a digital signal processor (DSP). Custom software was written for the circuit board that included the DSP to compute estimates of the spectral or Fourier components of observed current waveforms in real time. The board provided the DSP chip, a 16-bit, 2-channel analog-to-digital converter (ADC), a 16-bit, 2-channel digital-to-analog converter (DAC), off-chip memory for temporary storage, an off-chip flash PROM for permanent storage, and a convenient program-debugging interface. The DSP communicated data to the host PC over a parallel port. The host PC processed the spectral information to provide load recognition and other value-added services such as critical load diagnostics. Because this prototype was PC-based, it was possible to provide these services and deliver information remotely.

The computers labeled remote 1 and remote 2 in Figure 12 operated at 200 MHz and were equipped with 32 MB memory and approximately 1 GB hard drives. Both machines were equipped with a network interface card, a data acquisition card, and the DSP card. The computers ran a highly stable, open-source, free operating system and were remotely maintained by secure shell over the Internet. Stability is a key consideration for any system that must operate remotely in the field. At the test site, both systems operated without any direct intervention for over two years of constant use. During active testing, data collected over 24-hour periods were accumulated and compressed at the on-site computers and were automatically transmitted at night to laboratory computers, where NILM data were maintained on an FTP site for retrieval and analysis.

The NILM was conceived as a means of reducing the cost of obtaining electrical-power data, and it is worth making a rough assessment of costs and benefits relative to currently available metering technologies. The costs can be compared with conventional AC watt transducers, which require a voltage tap and current transducers on the input side and produce as output a low-voltage or current signal proportional to power. Both the NILM and a watt transducer require a personal computer to collect data. This computer can stand alone as an independent data logger or can be the same personal computer used for an energy management system. Both the NILM and a watt transducer require an A/D board to digitize analog information for the PC. This information is voltage and current data for the NILM and the power data for the watt transducer. Both the NILM and the watt transducer require current transducers and a voltage tap. The NILM can use the same PC, A/D board, current transducers, and voltage tap as the watt transducer. The NILM eliminates the watt transducer itself, performing the convolution of current and voltage in software, but requires an inexpensive voltage transducer ($25 for a research-grade transducer and a fraction of that for a transducer of adequate bandwidth) to lower the line voltage to a level appropriate for the A/D converter. The cost of the NILM is therefore comparable to a submeter, even if the goal is to monitor a single load. Multiple loads require multiple watt transducers, each with current transducers. A single-phase watt transducer with current transducer costs about $250-300. There were six loads of prime interest at the motor-control center in the ASHRAE test building: three supply fans and three chilled-water pumps. The costs are

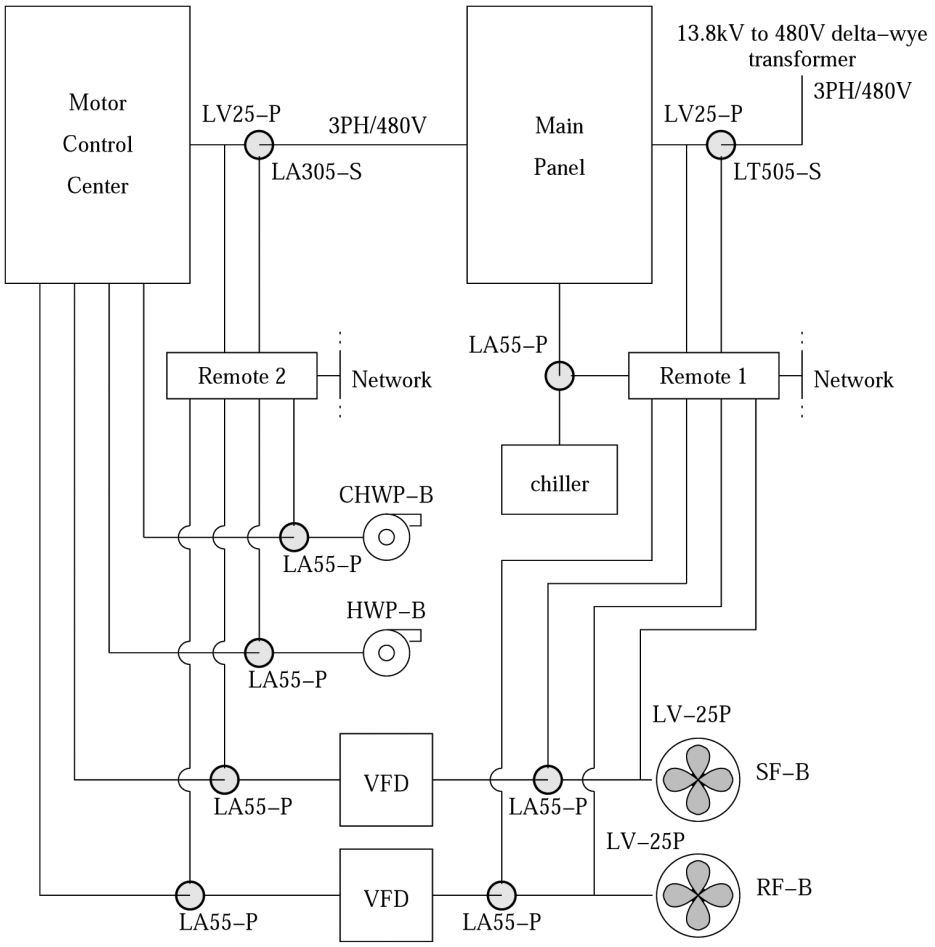


Figure 12 *Electrical schematic showing location of the two NILM meters at the test building. Sensors labeled LA55-P, LA305-P, and LT505-S are 50A, 300A, and 500A Hall-effect current sensors. LV25-P is an isolated voltage sensor.*

comparable for the first fan and the NILM saves the $1250-1500 required for the remaining two fans and three pumps.

More testing is required to determine whether the NILM can consistently produce power data of accuracy suitable for its intended use. Until these tests are done, submetered data remain the standard. But the NILM should not be evaluated as a trade-off between cost and measurement accuracy. It is a flexible platform that generates a data stream far richer than a submeter produces and that can analyze that data in ways limited only by the imagination of an engineer and skill of a programmer. For example, the power harmonics it calculates can be used for power-quality assessments and for identifying loads that generate certain harmonics. It can analyze very rapid start-up transients (Shaw et al. 2002) as a means of detecting faults in equipment performance and can detect power oscillations that are associated with actuator wear.

CONCLUSIONS

In principle, low-cost information about individual electrical components in a building can be obtained via careful analysis of power measurements at central locations within the building, notably the electrical service entrance and motor-control centers that supply power to HVAC components. Visual analysis of electrical power sampled at appropriate speeds shows step changes that can be associated with equipment of interest. The development of a reliable automatic detector suitable for use in noisy and complex electrical environments has involved selection of the basic statistical approach, development of guidelines for tuning the detector, and innovations that improve performance (enhanced sensitivity to signals of interest and rejection of electrical noise). Further work has extended the detection approach to identify sustained power oscillations, indicative of poorly tuned controllers, as well as step changes, and to operate on data

sampled at multiple rates, in order to fill in missing events and reject false alarms that may be generated from a single data set. Test results show that the centralized or non-intrusive load monitor can detect all on-off events of interest in some data sets from real buildings and that further work is required to automate the process of tuning detector parameters. With other data sets, the detector detected most, but not all, events. Further testing in other buildings is required, particularly to evaluate and automate the steps needed to tune the detector. In recent years of development, the detector's capabilities have been enhanced while the hardware cost has decreased, encouraging future development.

ACKNOWLEDGMENT

The authors warmly acknowledge the financial and technical support of Honeywell, Inc., the California Energy Commission (via subcontracts from Lawrence Berkeley National Laboratory and Architectural Energy Corporation), and ASHRAE.

REFERENCES

Abler, C., R. Lepard, S. Shaw, D. Luo, S. Leeb, and L. Norford. 1998. Instrumentation for high-performance nonintrusive electrical load monitoring. *ASME J. Solar Energy Engineering*.

Basseville, M., and I. Nikiforov. 1993. *Detection of abrupt changes theory and application*. Prentice Hall Information and System Sciences Series. Englewood Cliffs, N.J.: Prentice Hall.

Deschizeau, M., P. Bertrand, A. Anglade, and M. Grimaldi. 2000. A new method for detailed electric consumption of domestic appliances. *Proceedings of 2d international conference on energy efficiency in households appliances and lighting, Naples, Italy.*

Hart, G.W. 1992. Nonintrusive appliance load monitoring. *Proceedings of the IEEE* 80(12): 1870-1891.

Hill, R.O. 1995. Applied change of mean detection techniques for HVAC fault detection and diagnosis of power monitoring. M.S. thesis, MIT Department of Architecture.

Karl, W.C., S.B. Leeb, L.A. Jones, J.L. Kirtley, Jr., and G.C. Verghese. 1992. Applications of rank-based filters in power electronics. *IEEE Transactions on Power Electronics* 7(3): 437-444.

Leeb, S.B. 1993. A conjoint pattern recognition approach to nonintrusive load monitoring. Ph.D. thesis, MIT Department of Electrical Engineering and Computer Science.

Leeb, S.B., S.R. Shaw, and J.L. Kirtley. 1995. Transient event detection in spectral envelope estimates for nonintrusive load monitoring. *IEEE Transactions on Power Delivery.*

Leeb, S.B., and J.L. Kirtley. 1996. A transient event detector for nonintrusive load monitoring. Technical report. U.S. Patent 5,483,153.

Leeb, S.B., U.A. Kahn, and S.R. Shaw. 1998. Multiprocessing transient event detector for use in a nonintrusive electrical load monitoring system. Technical report. U.S. Patent 5,717,325.

Norford, L.K., and N. Mabey. 1992. Nonintrusive electric load monitoring in commercial buildings. *Proceedings of the eighth symposium on improving building systems in hot and humid climates*. Dallas, Texas.

Norford, L.K., and S.B. Leeb. 1996. Nonintrusive electrical load monitoring. *Energy and Buildings* 24: 51-64.

Norford, L.K., J.A. Wright, R. Buswell, D. Luo, S.R. Shaw, and S.B. Leeb. 2000. Demonstration of fault detection and diagnosis methods in a real building, ASHRAE RP-1020.

Norford, L.K., J.A. Wright, R. Buswell, D. Luo, S.R. Shaw, and S.B. Leeb. 2002. Demonstration of fault detection and diagnosis methods in air-handling units, ASHRAE RP-102. To be published in *International Journal of HVAC&R Research*.

Rice, J.A. 1988. *Mathematical statistics and data analysis*. Pacific Grove, Calif.: Wadsworth & Brooks/Cold Advanced Books & Software.

Shaw, S.R., D. Luo, L.K. Norford and S.B. Leeb. 2002. Fault detection via electrical load monitoring. To be published in *International Journal of HVAC&R Research*.

DISCUSSION

Rick Danks, Masa Glenn Research Center, Cleveland, Ohio: 1) Was the power distribution checked for integrity to avoid false readings due to loose connections, etc.? 2) Can this system distinguish between system problems (some listed above) and equipment problems?

Leslie K. Norford: The power-distribution system in the test building was very well maintained and of recent vintage. It was inspected visually during the installation of the two centrally located power monitors but was not subjected to other tests. The monitors are capable of detecting a side variety of problems but must be programmed accordingly. A reasonable analogy is that of a spot light that must be pointed in the right direction to reveal an object of interest, in contrast with a flood light that requires less tuning but reveals less detail. The monitors can be used to detect and diagnose such system problems as voltage sags due to start-up of large components, which may cause other equipment to abruptly shut down. With one-time measurements of impedances in the building electrical system, the monitor can estimate the voltage distortion at any point in the building. At the equipment level, the monitor can detect faulty electrical contactors.

Grant Wichenko, Appin Associates, Winnipeg, Canada: 1) Have you used power quality monitoring? 2) What can be determined with you system that you cannot get with a normal DDC system? 3) What resolution do you need to make sense of the data?

Norford: The electrical monitoring system we describe takes current and voltage measurements at high speed, calculates

harmonics of real and reactive power, and is capable of power-quality monitoring. A normal Building Energy Management System (BEMS) is not used for detection and diagnosis of HVAC faults. Extensive recent research, inside and outside ASHRAE, is producing FDD techniques that could be included in a BEMS. Our approach makes use of electrical-power data and could be used in lieu of or in addition to approaches that make use of thermofluid data (temperatures and flows). Further, our system is capable of finding some faults, such as misalignment in a motor-belt fan system, that thermal modeling would probably not detect.

Equipment-specific energy consumption is also important to building operators. In a conventional system, it is necessary to install expensive watt meters on each component of interest, an approach that is rarely taken due to cost. We are conducting field tests to determine whether we can provide an acceptable estimate of equipment-specific power via high-speed monitoring from a single point, a less expensive alternative.

The resolution required to make sense of the data depends on the application. We sample the current 128 times per line cycle, or 7,680 Hz. From these current data, we calculate real and reactive power and the fundamental and higher harmonics and produce a data stream at 120 Hz. We have down-sampled these data to obtain data at 1-10 Hz for FDD, based on step-changes in electrical power that indicate whether or not a pump, fan, or chiller has turned on or off. More fruitfully, we use the higher-speed data stream to resolve the start-up transients of motor-driven loads. These transients are very short, on the order of a second.

California Energy Commission
Public Interest Energy Research Program

HPCBS # E5P22T3b

HPCBS

High Performance Commercial Building Systems

Detection and Diagnosis of HVAC Faults via Electrical Load Monitoring

Element 5 - Integrated Commissioning and Diagnostics
Project 2.2 - Monitoring and Commissioning of Existing Buildings

S.R. Shaw, Montana State University
L.K. Norford and S.B. Leeb, Massachusetts Institute of Technology
D. Luo, United Technologies Corp.

VOL. 8, NO. 1
HVAC&R RESEARCH
January, 2002

pier
BERKELEY LAB
United States Department of Energy
Science, Security and Energy: Powering the 21st Century

Acknowledgement

This work was supported by the California Energy Commission, Public Interest Energy Research Program, under Contract No. 400-99-012 and by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

VOL. 8, NO. 1 HVAC&R RESEARCH JANUARY 2002

Detection and Diagnosis of HVAC Faults via Electrical Load Monitoring

S.R. Shaw, Ph.D. Member ASHRAE **L.K. Norford, Ph.D.** **D. Luo, Ph.D.** Student Member ASHRAE **S.B. Leeb, Ph.D.**

Detection and diagnosis of faults (FDD) in HVAC equipment have typically relied on measurements of variables available to a control system, including temperatures, flows, pressures, and actuator control signals. Electrical power at the level of a fan, pump, or chiller has been generally ignored because power meters are rarely installed at individual loads. This paper presents two techniques for using electrical power data for detecting and diagnosing a number of faults in air-handling units. The results from the two techniques are compared and the situation for which each is applicable is assessed.

One technique relies on gray-box correlations of electrical power with such exogenous variables as airflow or motor speed. This technique has been implemented with short-term average electrical power measured by dedicated submeters. With somewhat reduced resolution, it has also been implemented with a high-speed, centralized power meter that provides component-specific power information via analysis of the step changes in power that occur when a given device turns on or off. This technique was developed to detect and diagnose a limited number of air handler faults and is shown to work well with data taken from a test building. A detailed evaluation of the method is presented in the companion paper, which documents the results of a series of semiblind tests.

The second technique relies on physical models of the electromechanical dynamics that occur immediately after a motor is turned on. This technique has been demonstrated with submetered data for a pump and for a fan. Tests showed that several faults could be successfully detected from motor startup data alone. While the method relies solely on generally stable and accurate voltage and current sensors, thereby avoiding problems with flow and temperature sensors used in other fault detection methods, it requires electrical data taken directly at the motor, downstream of variable-speed drives, where current sensors would not be installed for control or load-monitoring purposes.

INTRODUCTION

The performance of many HVAC systems is limited more by poor installation, commissioning and maintenance than by poor design (Tong 1989). Computer-based control systems have the capability to collect and store sensor and control signals that could be analyzed to detect and diagnose faults. A considerable amount of research work has been carried out to develop FDD techniques for HVAC systems and, recently, to test these techniques in realistic laboratory settings and in real buildings (Ahn et al. 2001; Chen and Braun 2001; Dexter and Benouarets 1996; Dexter and Ngo 2001; Haves et al. 1996a; House et al. 2001; Hyvarinen 1996; Lee et al. 1996a, 1996b; Li et al. 1996; Peitsman and Bakker 1996; Salsbury 1996; Stylianou and Nikanpour 1996; Tsutsui and Kamimura 1996; Yoshida et al. 1996).

Steven R. Shaw is an assistant professor at Montana State University, **Leslie K. Norford** and **Steven B. Leeb** are associate professors with the Massachusetts Institute of Technology, and **Dong Luo** is a senior engineer with United Technologies Corporation.

A fault is detected when the observed behavior of a system differs from the expected behavior by some threshold. The expected behavior of the system is often expressed in a model, whether physical, statistical or fuzzy. The number of measurements required to use a given model for fault detection is of central importance, as there is a cost associated with installation and maintenance of sensors that are not required by the control system.

A fault is diagnosed when it is detected and a cause is determined. Diagnosis is significantly more difficult than detection, because measured effects must be attributed to a particular cause to the exclusion of all other possibilities. The difficulty of diagnosis increases with the number of possible causes, and careful validation in the context of the target application is required for any proposed diagnostic technique.

Current approaches to HVAC FDD have been driven by the confidence researchers have in their own modeling approach—its ease of implementation, robustness, and ability to generalize to many different HVAC faults—and in the availability and accuracy of measured data. The approach in this paper begins with the premise that electrical power measurements are useful in FDD, based on the authors' experience extracting device-level information from centralized power measurements. The paper presents FDD techniques using submetered power data that, in some cases, may be readily adapted for use with centralized measurements. Whether these techniques are economically and technically useful needs to be explored in the field over the coming years.

There are four generic approaches to consider for use of electrical power data in HVAC FDD, as shown in Table 1. Power signals can be obtained from meters attached to fans, pumps, chillers, or other individual pieces of electrically powered HVAC equipment. Alternatively, power data can be collected at a single point. Further, detection and diagnosis of faults can be based on analysis of changes in steady-state electrical power, or analysis of the dynamic variation of power during the typically very brief startup transient. For submetered power, steady-state is defined as one-minute averages at times other than those encompassing a startup or shutdown event or substantial power oscillation. In contrast, for centralized power measurements, steady-state power is the difference in power before and after a turn-on or turn-off event. Techniques presented in this paper cover three of the four approaches, as noted in Table 1. The detection and diagnosis of faults from startup transients recorded centrally, the most powerful but difficult method, has not been attempted to date.

Centralized load monitoring takes its name of Non-Intrusive Load Monitoring (NILM) from its origins in load analysis in houses, where the normal revenue meter was replaced by a meter capable of clustering step changes in real and reactive power and associating these clusters with major household appliances (Hart 1992). Over the last decade, the early techniques have been extended in two directions. First, enhanced detection of step changes in power have been shown to screen out power spikes caused by switching electronics and to find power oscillations caused by poorly tuned controllers (Norford and Mabey 1992, Hill 1995, Luo et al. 2002). Second, a powerful detection approach has been developed to analyze the rapid changes in power that occur when a motor, lamp, or computer is first turned on (Leeb 1993, Leeb et al. 1995, Norford and Leeb 1996, Abler et al. 1998). The startup patterns are governed by the physics of the device and are generally not masked by power electronics that reduce reactive power. In addition, anal-

Table 1. Options for Detecting and Diagnosing HVAC Faults by Analyzing Electrical Power Signals Explored in this Paper

Location of Meters	Analysis of Changes in Steady-State Power	Analysis of Power Dynamics During Equipment Startup
Individual loads	X	X
Centralized	X	

ysis of startup transients is useful in a busy electrical environment, due to the very brief amount of time (typically fractions of a second) required to log a characteristic signal. Both approaches to centralized load monitoring rely on a combined hardware/software platform that is considerably more capable than the typical Watt meter and of comparable cost. This platform has been refined to the extent that it consists solely of the necessary transducers fed into a personal computer for analysis and communication via the internet.

This paper focuses on the use of electrical power signals for HVAC FDD and not on centralized load monitoring. One major motivation for this work has been an ASHRAE-sponsored research project, 1020-RP, "Demonstration of Fault Detection and Diagnosis Methods in a Real Building." In this project, one electrical power FDD approach, using gray-box correlations of steady-state changes in electrical power with flow or other variables, was tested on three variable air volume (VAV) air-handling units (AHUs) in a heavily instrumented test building. A second FDD approach (not required for 1020-RP) was also tested with data from the test building. This approach uses a physical model of the dynamics of a motor and driven load during a startup transient. Parameters estimated solely from electrical current measurements are compared with parameters estimated during periods of normal behavior to detect faults.

The HVAC system, diagrammed and described more fully in the companion paper, Norford et al. (2002), includes variable-speed supply and return fans, constant supply-air and chilled water temperature control, and a primary-secondary pumping system with constant-speed secondary pumps. The seven AHU faults included in these tests, listed in Table 2, are featured in this paper. Table 2 notes whether the selected faults were abrupt or occurred slowly over time; as implemented in the test building, abrupt faults were introduced as such and degradation faults were introduced over one- to three-day periods. Electrical power FDD methods are no different from others in their ability to find abrupt faults more easily than degradations.

Test results presented in this paper are limited to the faults introduced in the test building and as such are demonstrations of the methods rather than comprehensive assessments of their efficacy. A final report and the companion paper (Norford et al. 2000, 2002) summarize the results of the blind tests conducted as part of 1020-RP. This paper, in effect, lays a foundation for the summary paper.

While the presentation focuses on a small number of artificially introduced faults, the presented FDD methods can in principle be extended to cover additional AHU faults and faults in other systems. The obvious prerequisite is that any fault to be detected by these methods must

Table 2. List of Air-Handling Unit Faults Detected and Diagnosed with Electrical Power Data

Fault	Type
Air Mixing Section	
Stuck-closed recirculation damper	Abrupt
Leaking recirculation damper	Degradation
Filter-Coil Section	
Leaking cooling coil valve	Degradation
Reduced coil capacity (water-side)	Degradation
Fan	
Drifting pressure sensor	Degradation
Unstable supply fan controller	Abrupt
Slipping supply fan belt	Degradation

Note: Fault implementation is described in Norford et al. (2002).

cause a change in electrical power. Further, the power change must be sufficiently large to cause a noticeable deviation from expected behavior, which in turn must be defined on the basis of gray-box power correlations using steady-state data or a physical model of a motor's startup transient. The gray-box power correlations require an understanding of the physics of the system under investigation, to determine power correlations suitable for detecting a given fault. The method therefore requires a considerable amount of expert knowledge to be made more generally applicable.

APPROACH

A. Detection and Diagnosis of HVAC Faults with Gray-Box Models and Submetered Electrical Power Measurements Under Steady-State Conditions

The gray-box electrical power method for fault detection and diagnosis rests on the following four steps:

1. A training phase, consisting of correlation of electrical power as measured under steady-state conditions with load measurements (gray-box models)
2. Measurement of electrical power
3. Detection of faults by comparison of measured and predicted electrical power and by screening for rapid power oscillations
4. Diagnosis of faults by rule-based analysis of power deviations

The gray-box method requires that fan, pump, and chiller power be correlated with an indicator of load for each electrical component. These correlations establish models for detecting equipment faults. For equipment such as that installed in the test building, Tables 3 and 4 summarize how these correlations depend on building thermal loads, as expressed by outside temperature in each of four regions:

1. $T_{out} < T_{balance\ point}$
2. $T_{balance\ point} < T_{out} < T_{supply\ air} - \Delta T_{supply\ fan}$
3. $T_{supply\ air} - \Delta T_{supply\ fan} < T_{out} < T_{return\ air}$
4. $T_{return\ air} < T_{out}$

The balance point temperature is the outside dry-bulb temperature at which the building requires neither heating nor cooling. In commercial buildings, $T_{balance\ point}$ accounts for a minimum intake of outdoor air required to satisfy ASHRAE *Standard* 62. Above $T_{balance\ point}$ there is a net cooling load and the supply airflow will increase. In some cases, $T_{balance\ point}$ may exceed $T_{supply\ air}$, in which case the two are simply exchanged in the temperature segments. The supply air temperature is assumed to be constant, as was the case in the test building. In Table 4, the coil capacity and slipping fan belt faults are best detected at high cooling loads but can be detected at lower loads if sufficiently severe.

Correlations of Fan Power with Airflow

If electrical power measurements for HVAC equipment are used as the basis for an FDD method, there must be some model of power under known load conditions. Studies have established that whole-building fuel use correlates with outdoor temperature (Fels 1986, Ruch and Claridge 1992). At the device level, fan power is of key importance in detecting many AHU faults. Previous work (Lorenzetti and Norford 1992) showed that hourly average power measurements for VAV supply and return fans correlate well with outdoor dry-bulb temperature. A simple three-variable change-point model was found to work as well as or better than other

Table 3. Variation of Outside and Supply Airflows and Pump, Chiller, and Fan Powers with T_{out}[a]

Temperature Region and Control Mode	Outside Airflow	Supply Airflow	Pump Power	Chiller Power	Fan Power
1 Heating	Constant at minimum value	Constant at minimum value	Off	Off	Constant at minimum value
2 Free cooling via air-side economizer	Increasing with T_{out}	Increasing with T_{out}	On; constant power	Energized; minimal power to meet thermal loads in piping	Increasing as a polynomial function of flow or speed
3 Mechanical cooling with 100% outside airflow	Increasing with T_{out}	Increasing with T_{out}	Nearly constant power if piping is balanced	Increasing with T_{out}	Increasing as a polynomial function of flow or speed
4 Mechanical cooling with minimum outside airflow	Constant at minimum value	Increasing with T_{out}	Nearly constant power if piping is balanced	Increasing with T_{out}	Increasing as a polynomial function of flow or speed

[a]For a VAV system with variable-speed supply and return fans, constant supply air temperature, and a constant-speed secondary chilled water pump.

Table 4. Detectability of Faults with Electrical Power Data

Temperature Region	Stuck-Closed Recirculation Damper	Leaky Recirculation Damper	Leaky Cooling Coil Valve	Reduced Cooling Coil Capacity	Pressure Sensor Drift	Unstable Pressure Controller	Slipping Supply Fan Belt
1	Yes	No	No	No	Yes	Yes	Maybe
2	No	No	Yes	No	Yes	Yes	Maybe
3	No	Yes	No	Maybe	Yes	Yes	Maybe
4	Yes	No	No	Yes	Yes	Yes	Yes

models fitted to these data. Outdoor dry-bulb temperature is a reasonable predictor because the sensible fraction of envelope loads (latent heat transfer excluded) influences the total amount of air delivered to occupied spaces. However, this predictor does not account for variations in airflow, and hence fan power, due to changes in internal or solar loads. Correlations with measured airflow provide a more precise estimate of fan power, as established in an earlier study of applications of electrical load monitoring to fault detection in ventilation systems (Norford and Little 1993). Such correlations have also been used to estimate fan energy consumption before and after variable-speed drive retrofits (Englander and Norford 1992).

Figure 1 shows a third-order polynomial correlation between fan power and airflow for a VAV supply fan with a variable-speed drive. The use of a third-order polynomial correlation is based on the fan laws, which show that power varies as a cubic function of speed for a variable-speed centrifugal fan. A similar correlation has also provided a good fit in practice for data collected from VAV fans equipped with inlet vanes. Ninety percent confidence intervals were established from uncertainties in the polynomial coefficients and a *t*-statistic. The confidence intervals express the confidence that a single new measurement point will lie between the upper and lower intervals, if the measurement is subject to the same conditions as occurred during the training phase. Increasing the confidence interval would make the method less sensitive to faults and less likely to generate false alarms. In the test building, the use of 90% confidence intervals did not generate false alarms, but in practice the confidence interval could be enlarged, reducing the number of both detected faults and false alarms. To tighten the correlation and improve the sensitivity of the method, only training data with duct static pressure within 25 Pa (0.1 in. of water) of the 300 Pa (1.2 in. of water) set point were accepted. This tolerance on duct static pressure was arbitrarily selected after the data were examined by eye and was intended to elimi-

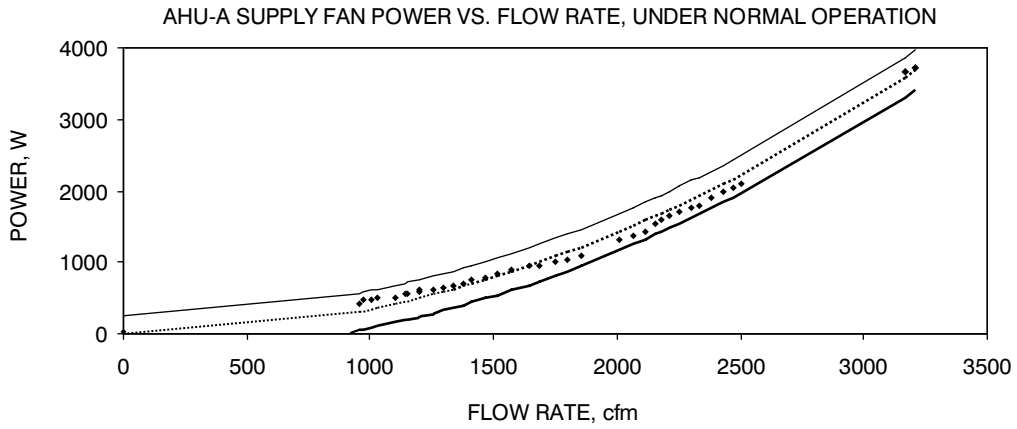


Figure 1. Correlation of Fan Power with Airflow During a Training Period

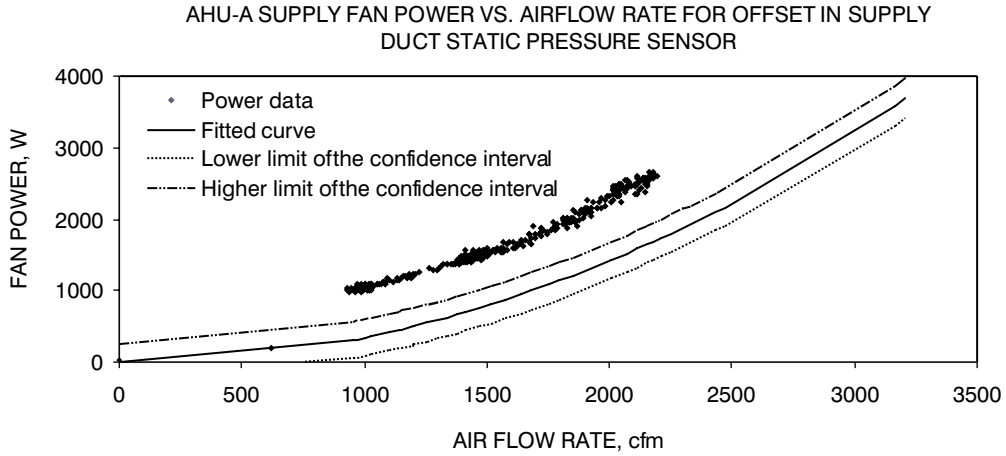


Figure 2. Fan Power Data Outside 90% Confidence Interval Established with Normal Operating Data, Due to Offset of 150 Pa (0.2 in. of water) in Supply Duct Pressure Sensor

nate data points collected during system transients when the supply fan speed was changing to bring the duct static pressure back to the set point.

This method has the desirable property that the threshold for fault detection is determined solely by the selection of a desired confidence interval and the distribution of data during the training phase. For example, unevenly distributed data sets will cause the confidence intervals to expand in airflow regions where data are scarce (Norford and Little 1993). The threshold for fault detection at any flow is therefore not arbitrary but is statistically defensible, under the major assumption that data collected during the training phase represent normal operation of the AHU.

Correlations of power with flow can be used to detect faults involving duct pressure, including pressure sensor faults, stuck dampers, and fouled filters. Figure 2 shows data well outside the confidence intervals, obtained in a fault condition when air was bled from the static pressure

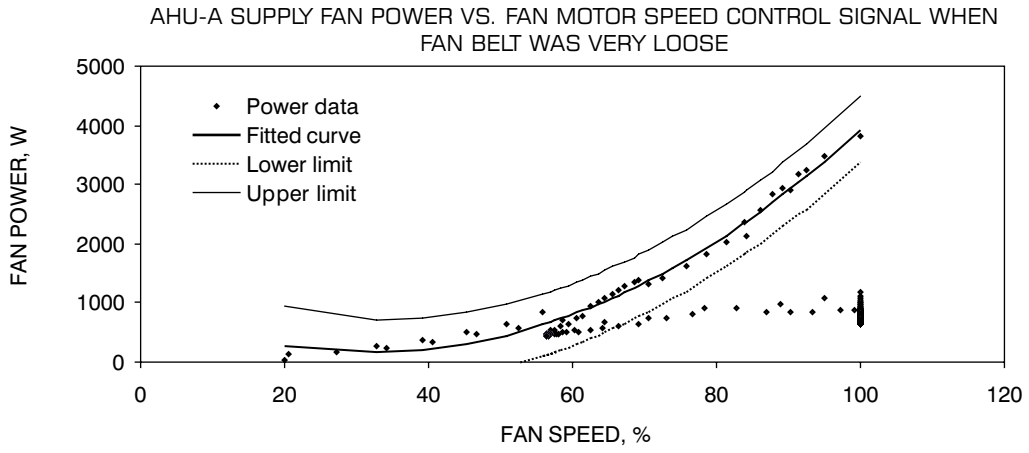


Figure 3. Reduction of Fan Power at Very High Speed-Control Signals, Due to Slipping Fan Belt

sensor's pneumatic line to simulate an offset in the static pressure sensor. The offset was introduced in three stages of 50, 100, and 150 Pa (0.2, 0.4, and 0.6 in. of water), with the most extreme case illustrated in Figure 2.

The correlation shown in Figures 1 and 2 can be used to detect a pressure sensor offset, either positive or negative, at any time of year. The same correlation can also be used to detect the stuck-closed recirculation-damper fault, but only under conditions identified in Table 4, where the damper would normally be wide open or nearly so. Under such conditions, a stuck-closed damper will increase the pressure drop in the air handler and therefore increase fan power for a given airflow.

Correlations of Fan Power with Fan-Motor Speed-Control Signal

Some faults can be detected when power is correlated with motor speed control signal, rather than airflow. For example, reduced tension in a fan belt can cause the belt to slip and transmit less mechanical energy from the motor to the fan at high load. If there is relatively little dissipation of energy in the loose belt, then, for a given motor speed, the motor draws less power and the airflow is lower. Fan motor power is not expected to change significantly for a given airflow. However, the power drawn by the motor for a given rotational speed will be less than normal, because the fan is not spinning as fast as it normally would and therefore transfers less power to the air. This fault can therefore be detected by a deviation from the expected relationship between fan power and the motor speed control signal, as shown in Figure 3. It is important to note that the power-speed correlation is immune to changes in duct pressures that can be detected with the power-flow correlation, as was established in Norford and Little (1993). As a result, the method for detecting the slipping fan belt is also sufficient to distinguish it from the faults detected with the power-flow correlation. However, a power-speed correlation will detect such additional faults as a loose pulley or a transducer failure, making fault diagnosis problematic. Figure 4 shows a transducer failure implemented in the test building.

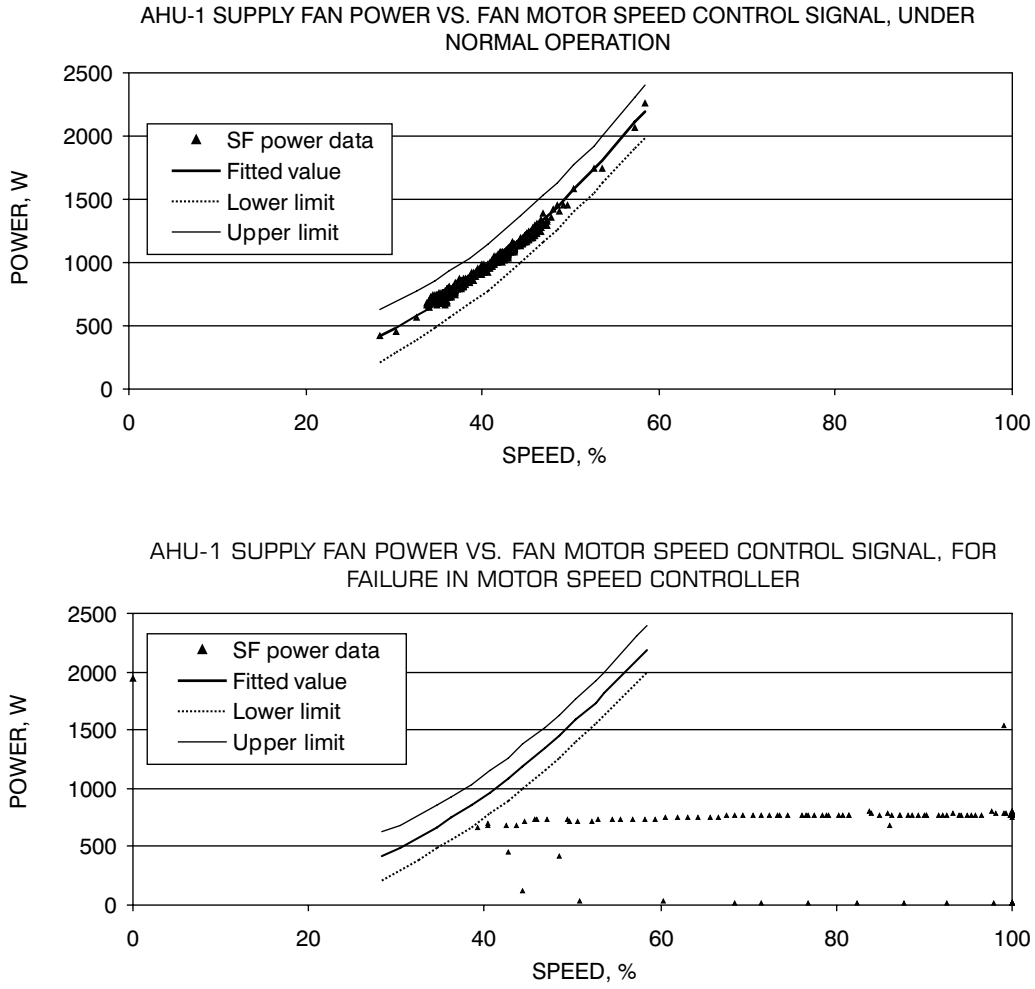


Figure 4. Reduction of Fan Power at High Values of Speed-Control Signal, as Result of Controller Failure

The top figure shows the fit to data from a single day of normal operation. The lower figure shows data during a three-hour period when the fan speed was fixed at a single value, over a range of speed-control inputs.

Correlation of Chilled Water Pump Power with Cooling Coil Valve Position

Changes in pump power can be used to detect blockages in piping, in the same manner as using fan power to detect stuck dampers. In many buildings with a single chilled water loop, the chilled water pump is run at constant speed and flow to the cooling coil is controlled with a three-way valve. The pump will ride the pump curve as flow resistance changes and pump power will change accordingly. Flow resistance can change due to normal operation (operation of the three-way valve to control flow to the coil, which has a larger flow resistance than the by-pass piping unless the latter includes a balancing valve) or due to a fault (flow obstruction). With

a two-way valve in lieu of a three-way valve, pump power will vary more strongly with valve position. While changes in pump power are an attractive means of detecting flow-restriction faults, chilled water flow is rarely measured and therefore is not a desirable correlating variable. The one available control signal, which adjusts the position of the cooling coil valve (either two-way or three-way) was used.

The test building has a primary-secondary piping system, with a variable-speed primary pump and constant-speed secondary pumps for the two test air handlers. Initially, the cooling loops used two-way valves to control flow to the air-handler cooling coils, but these valves were replaced with pairs of two-way valves that are controlled to perform as three-way valves. Further, flow balancing valves were installed to equalize the pressure drop across the coil and the bypass piping. The pump power under normal operation is expected to be nearly independent of valve position. If there is a flow restriction in a cooling coil when balancing valves are installed and used properly, it is expected that there will be less flow through the coil; the discharge air temperature will tend to rise above its set point; the valve controller will open the valve to send more flow through the coil and less to the bypass loop; the overall flow resistance will increase and the pump will ride up the pump curve to a lower total flow and reduced pump power.

Correlating pump power with valve control signal is therefore sufficient to detect flow blockages under cooling loads sufficiently high that a substantial fraction of the total flow is directed through the coil. However, it is not likely to detect coil fouling, where a very thin coating of calcium carbonate can drastically reduce heat transfer across the coil but can have a small impact on flow resistance. The change in pump power as correlated with valve position is illustrated in Figure 5. The training period did not include system loads large enough to cause the cooling coil valve to open more than 70%. Pump power was assumed to remain nearly constant under higher flows through the cooling coil.

Chiller power can provide another indication of faults in chilled water piping. For low to moderate cooling loads, the valve controller will compensate for a flow restriction in the cooling coil by directing more water to the coil and less to the bypass piping. Under high cooling loads, the valve controller will saturate, flow through the coil will be less than needed to maintain the discharge-air temperature, the building will be undercooled, and chiller power will therefore drop. Monitoring to determine a reduction in chiller power at high load also offers the advantage of detecting not only flow-restriction faults but also reduced thermal conductivity due to deposits on the water side of the cooling coil.

Whole-building energy studies have correlated building electricity consumption with outside temperature as a means of analyzing the building's energy requirements for cooling, typically with linear change-point models (Ruch and Claridge 1992, Ruch et al. 1993). More detailed studies of chiller power have established that chiller power is primarily a function of load on the chiller and the temperature difference between leaving condenser water and chilled water flows, and that a biquadratic functional form is a reasonable model (Braun et al. 1987). For an air-cooled chiller as is used in the test building, outside air temperature directly affects condenser performance.

In this study, an HVACSIM+ simulation of a building modeled for a controls simulation test bed (Haves et al. 1996b, 1998) was used to correlate cooling load, as measured by heat transfer across the cooling coil, against outdoor temperature. It was possible to detect a fouled coil. This method was not applied in the test building to detect the coil capacity fault because the chiller is a two-stage reciprocating unit and power levels are discrete, rather than continuously varying. Cycling periods between states were not regular and not easily discerned at high loads. At low cooling loads, however, chiller cycling was both regular and revealing of certain faults, as will be discussed.

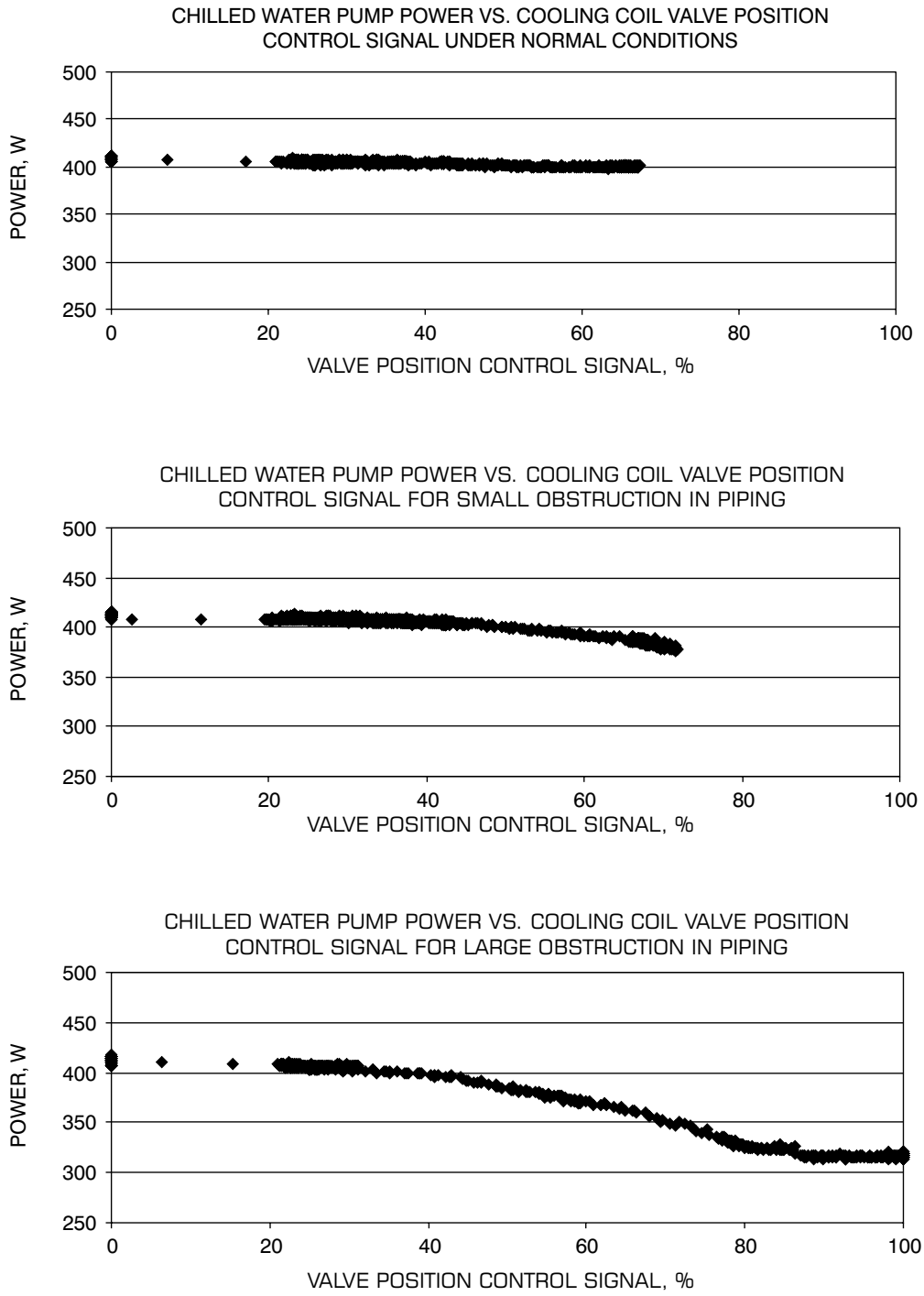


Figure 5. Correlation of Pump Power with Valve Position, for Normal Conditions and for Reduction in Cooling Coil Capacity Due to Obstruction in Chilled Water Piping

Analysis of Reciprocating Chiller Cycling Periods

Cycling periods for the two-stage reciprocating chiller, between shutdown and the low-power state, can be analyzed to detect such faults as a leaky recirculation damper and a leaky cooling coil valve. Both of these faults can be detected only at low cooling loads, when the recirculation damper would normally be shut by the air-side economizer and when the cooling coil valve would be normally closed. Under low loads, a reciprocating chiller will cycle on for a short period and then shut down for a longer period, controlled to maintain the chilled water temperature returning to the chiller within a specified band. For the leaky recirculation damper, the chiller cycling period can be analyzed in a narrow band of outside temperatures, just above the point where the cooling coil valve starts to open and there is a cooling load across the coil. (This corresponds to the boundary between outdoor temperature regions 2 and 3, as defined above.) Limiting the analysis window to a narrow band guards against labeling as a fault those changes in cycling period that are due to a normal increase in cooling load with outside temperature.

This approach works well when outdoor temperatures exceed the supply air temperature set point by a small fraction of the supply-return temperature difference (20% has been used as a cutoff), and, importantly, remain in this region long enough to detect a stable chiller cycling pattern. When the outdoor temperature rises and the building cooling load increases, as is the case on many days, chiller-cycling period is not a good fault detection metric.

The leaky cooling coil valve will increase the load on the chiller for a given valve position. It is difficult to correlate changes in reciprocating chiller cycling periods with valve position, due to the coarse nature of the former. An acceptable alternative is to analyze changes in chiller-cycling period when the valve control signal is zero and the valve is nominally shut. This approach is relatively robust, with less chance for false alarms than if chiller-cycling data were considered when the valve is partially open. It also distinguishes the symptoms of this fault from the leaky recirculation damper.

Figure 6 shows a chiller-cycling period of 20 to 25 min. when the cooling coil valve was closed, much shorter than the normal cycling period of 38 to 39 min. observed earlier the same day and indicative of a leaky valve. The cycling interval dropped to 14 to 15 min. when the cooling coil valve was about 20% open later in the day, a period excluded from analysis.

Analysis of Fan Power Oscillations

Norford and Leeb (1996) showed that centralized power monitoring could detect chiller-power oscillations due to an unstable chilled water temperature controller. Screening for rapid power oscillations forms the basis for detecting faults caused by underdamped or unstable local-loop controllers. Power oscillations are quantified by the standard deviation of the data sampled in a sliding window. In the test building, this approach has been used to detect an unstable pressure-control-loop fault in the test building, as shown in Figure 7.

B. Detection and Diagnosis of HVAC Faults with Gray-Box Models and Centralized Electrical Power Measurements Under Steady-State Conditions

The steady-state, gray-box FDD method was also evaluated when the electrical power data came from centrally located power meters rather than submeters dedicated to individual components. Two centralized meters, known as Non-Intrusive Load Monitors (NILMs), were installed in the test building, one on the motor control center serving five fans and ten pumps, and one at the whole-building electrical service entry. A detailed development of the signal-processing algorithms needed to detect steady-state changes in HVAC loads is presented in Luo et al. (2002). The use of the centralized meters to detect classes of faults, including six of the seven introduced in the test-building air handlers, is briefly described.

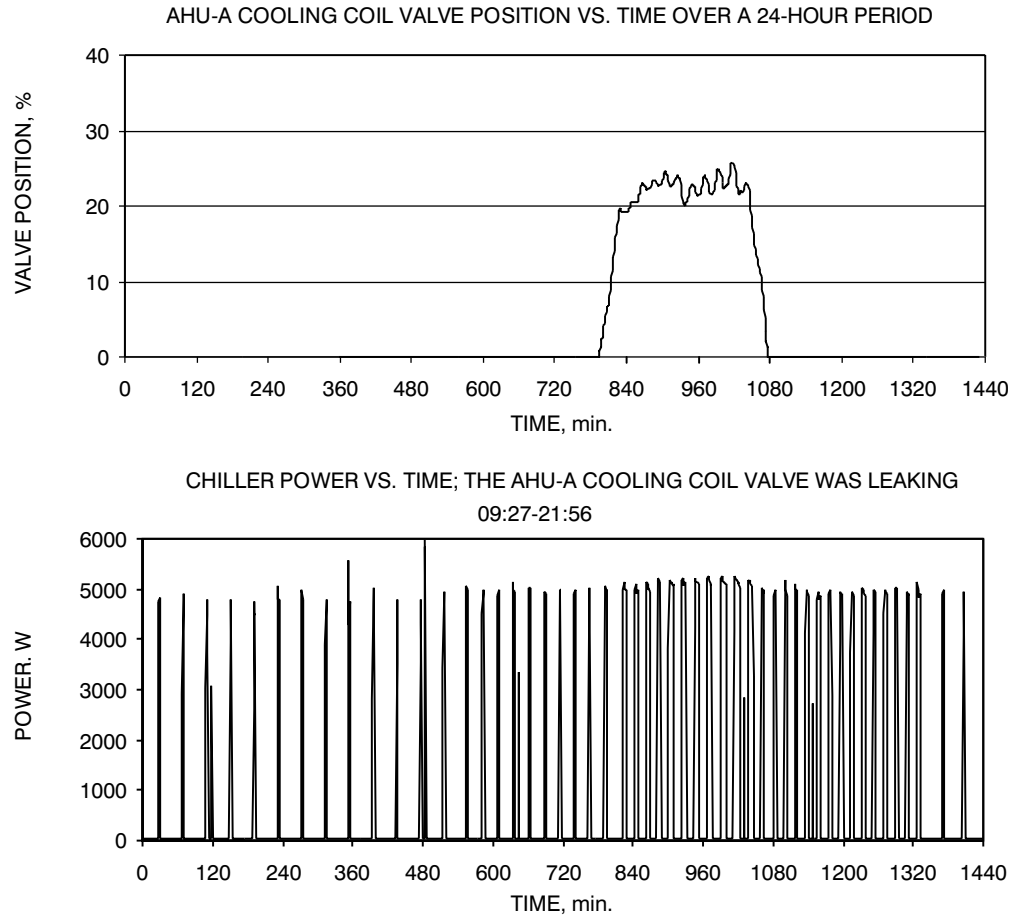


Figure 6. Detection of Leaky Cooling Coil Valve by Analysis of Cycling Period of Reciprocating Chiller

The power-flow correlation for the fan that was readily made with submetered power data could not be generated with electrical power measurements from the NILM installed on the motor control center in the test building. The NILM yielded one data point per day, when the fan was turned off in the evening. The startup point was not valuable because the fan motor has a variable-speed drive that has a slow and complex startup pattern that is very different from an abrupt change most easily seen by the change-of-mean detection algorithm applied to data collected by the NILM. Further, there was little variation in flow at the time of fan shutdown and no opportunity to generate a polynomial relationship between power and flow. The limited range of data also made it impossible to correlate fan power with the motor-speed-control signal, a correlation established with submetered data and used to detect and diagnose the slipping fan belt.

The fact that airflow, fan speed and fan power showed little variation at shutdown from day to day, due to nearly constant (and very light) building loads in late evening, made it possible to use deviation of fan power from the normal value as a basis for fault detection. For example, when the fan belt in the test building slipped severely, the fan power was about 750 W with a

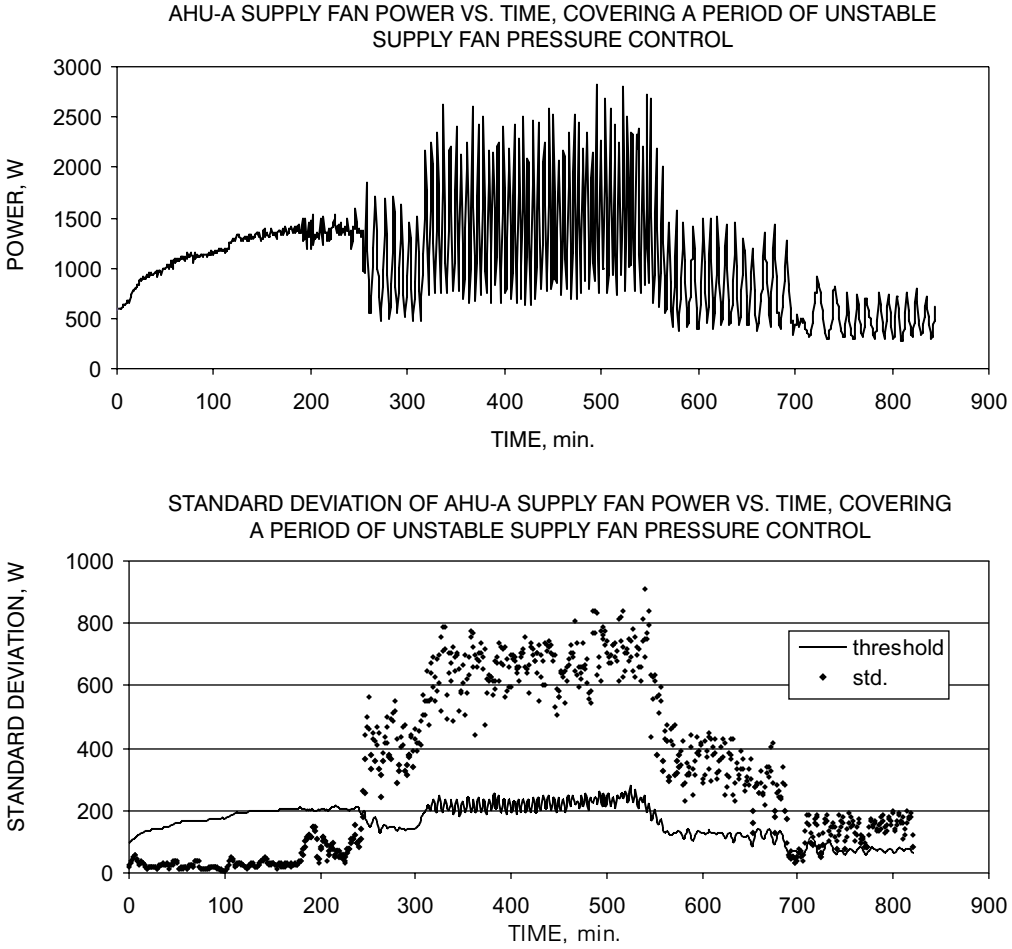


Figure 7. Detection of Unstable Pressure Controller via Analysis of Standard Deviation of Supply Fan Electrical Power

100% motor speed control signal at shutdown, compared to a normal speed of about 20% and a normal power of about 500 W.

Supply fan electrical power of 600 W at shutdown, 100 W above the typical value, was taken as a threshold above which a fault was flagged. This value, selected after examination of the data, proved acceptable in that it did not generate false alarms in the test building. Among the limited number of faults introduced into the test building, the three possible causes were the stuck-closed recirculation damper, an offset in the supply duct static pressure sensor, and the slipping fan belt. Power data alone were sufficient to detect these faults but not to diagnose them. Motor speed data would separate the slipping fan belt from the two faults that are driven by pressure changes.

Figure 8 shows fan power as estimated by the NILM and the supply fan submeter. Power is plotted against airflow to show the limited range of airflows at shutdown, compared with the operating range shown earlier in Figures 1 and 2. The NILM estimate of power is systematically lower than the submeter, a discrepancy that stems from the difference between one-minute

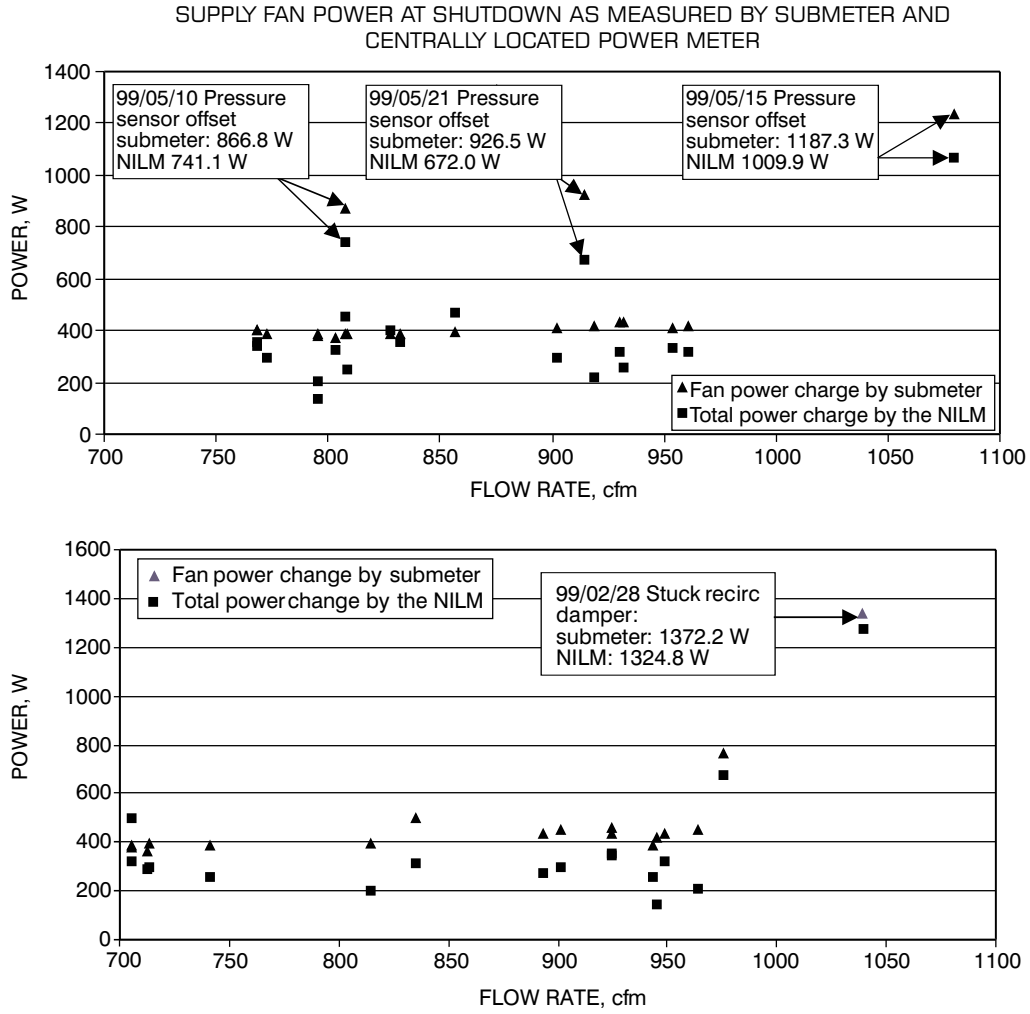


Figure 8. Detection of Pressure Sensor Offset (Top) and Stuck Closed Recirculation Damper (Bottom), via Analysis of Fan Power at Time of Shutdown, as Recorded by Centralized Power Meter with Fault Detection Threshold of 600 W

power averages as used in the submeter and the NILM's calculation of power on the basis of short-term data immediately before and after the shutdown event is detected.

There are two alternative approaches that avoid establishing a somewhat arbitrary power threshold for detecting faults at the time the fan is shutdown. First, fan power as a function of flow and static pressure set point can be modeled from manufacturer's data (Englander and Norford 1992). The model can be tuned with one or more data points. However, such a model does not provide the statistical confidence intervals that come from a polynomial fit of power to flow and a threshold for fault detection would still need to be assigned. Second, fan power data at times other than shutdown can be used to generate the polynomial relation between fan power and airflow or motor speed. Power data can be obtained with a portable power meter used during a commissioning period, or by shutting down the fan several times during a training period,

when airflows or speeds are at values that span the expected range of operations. As with submetered power data, confidence intervals establish a range over which data are considered to be normal and no arbitrary threshold is required. Power at shutdown can then be compared with the expected power for the flow measured at shutdown. This approach was implemented and the power estimate that came from fitting power to airflow during the training period was found to closely predict subsequent power measurements made with submeters.

Submetered power for the secondary chilled water pump was adequate to detect the coil-capacity fault. However, the pump power was so small that it was difficult to reliably detect from the NILM at the motor control center, without lowering the detection threshold to the point where small, unknown, and uninteresting loads were flagged. Further, successful detection of this fault requires accurate analysis of small changes in the small load. Therefore, the NILM was not used to detect pump power.

Luo et al. (2002) describe an effective means for using the meter at the building service entry to pinpoint chiller cycling, even in a very noisy electrical environment. This method boosts sensitivity and reduces false alarms by sampling the power signal at multiple rates, detecting on- and off-events for each data set, and combining the results to reconstruct the sequence of on- and off-events for a given component.

This method has been used to successfully detect both the leaky cooling coil valve and the leaky recirculation damper in the test building. As noted in the discussion of FDD with submeters, the two faults were distinguished by insisting that the cooling coil valve be closed as a prerequisite for diagnosing the leaky valve and by focusing on a narrow range of outdoor temperatures when the recirculation damper should normally be tightly shut and chiller loads are low. Figure 9 shows the use of the NILM at the building service entry to detect the leaky recirculation damper. The middle plot, based on submetered power data, shows a chiller cycling interval of about 15 min. over a 10-h period when the outdoor dry-bulb temperature was sufficiently low. The lower plot shows the output of the NILM. Positive changes in power are associated with the chiller's going from an off-condition to the lower of its two stages, and negative power changes signify shutdown events. The NILM detected all of the transitions, which is what was needed to detect the leaky recirculation damper, although the power the NILM associated with these transitions differs from the submetered value of 5 kW. The normal cycling period was observed to be well above the 30-min. threshold established for this fault.

Power oscillations were detected with the NILM attached to the motor control center, indicative of the unstable-supply-duct-pressure-controller fault introduced in the test building. In general, fault diagnosis from a central point is not possible. While it may be possible to assign the magnitude of the power oscillations to classes of equipment (fans, pumps, chillers), it is not possible to pinpoint a particular fan or pump if there are multiple units of comparable size.

C. Detection and Diagnosis of HVAC Faults with Dynamic Models and Submetered Electrical Power Measurements

The chilled water pump and the supply fan associated with AHU-B in the test building were considered for detection and potential diagnosis of faults using a dynamic-modeling approach. The method used involves fitting parameters to a physical model of the startup behavior of the motor and its attached load using only current measurements available at the stator of the motor. Changes in these physical parameters provide useful HVAC diagnostic information such as flow obstructions.

Pump Diagnostics

Deposits in the pipes of a heater exchanger are difficult to detect noninvasively and can contribute to decreased heating and cooling efficiency. In principle, it might be possible to detect

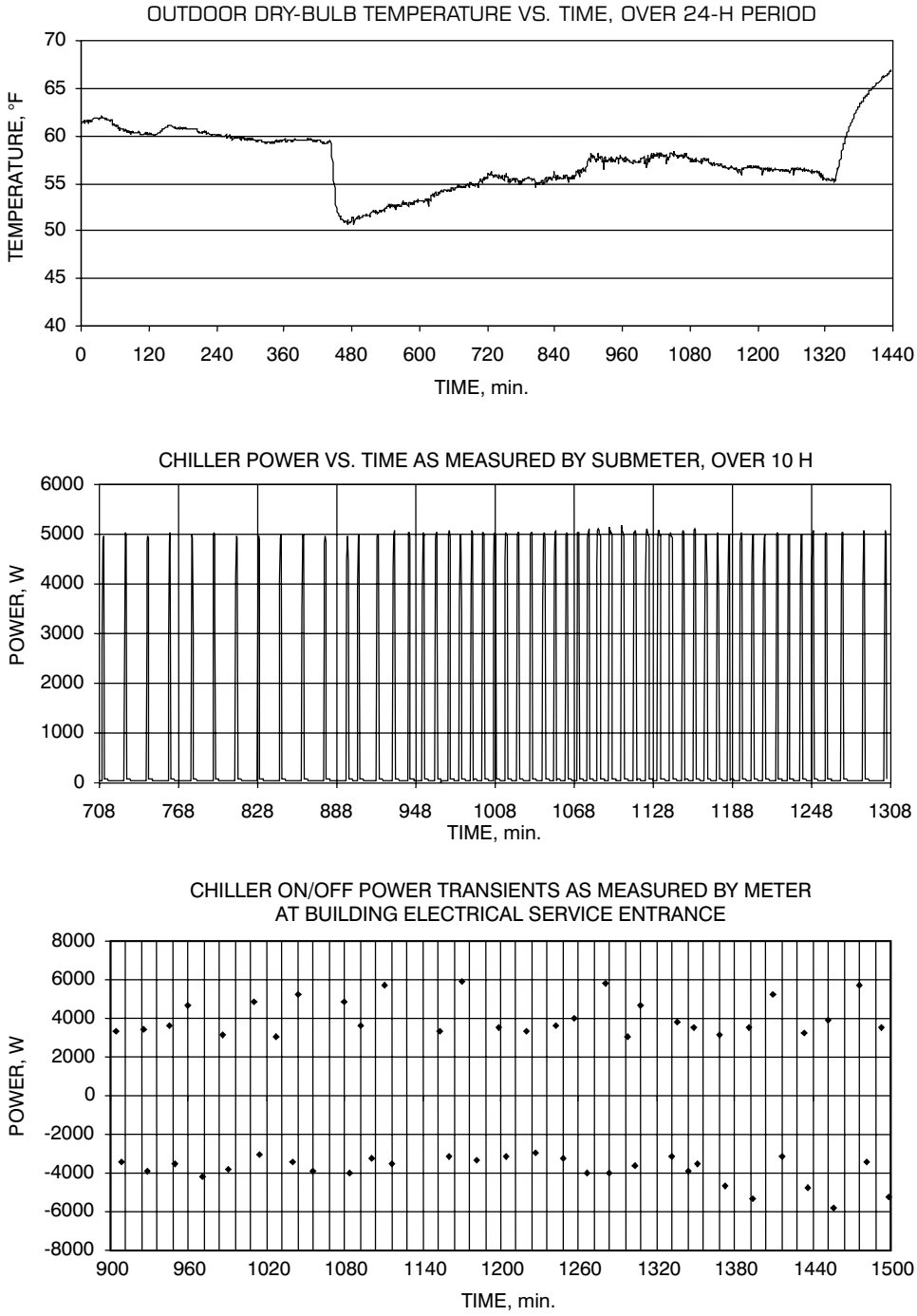


Figure 9. Detection of Leaky Recirculation Damper via Analysis of Reciprocating Chiller Cycling Intervals with Outdoor Temperature from 13 to 15°C (55 to 59°F)

The second plot shows submetered chiller power and the lower plot reveals the power changes (positive for turn-on, negative for turn-off) detected by the NILM.

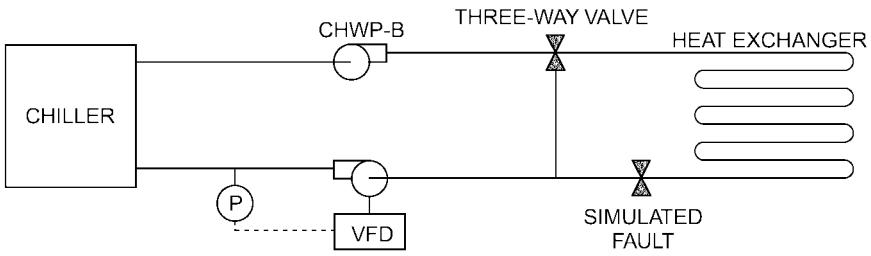


Figure 10. Simplified Diagram of Cooling Loop for AHU-B in Test Building
A single variable-speed primary pump and a fixed-speed secondary pump were used for each air handler. An obstructed cooling coil was simulated with a throttling valve. Current to the secondary pump CHWP-B was measured.

deposits by estimating flow resistance for the fluid in the pipes. Electrical transients were recorded for a chilled water pump in the HVAC system for both normal and an obstructed-flow condition. Detection of the obstructed flow is demonstrated via parameter estimation using the collected data.

A simplified diagram of the chilled water circulation system is shown in Figure 10. The chilled water system consists of a single, variable-speed primary pump and one fixed-speed secondary pump for each air handler. For the test, the three-way cooling coil valve was positioned so that all liquid flowed through the heat exchanger. The primary chilled water pump, equipped with a variable-frequency drive, was operated to control pressure as indicated. The response of this control loop was presumed to be slow enough to ignore in comparison to the startup transients of the secondary pump, CHWP-B. Further, the impact of the primary loop on the secondary pump startup transient could be accurately modeled by including the fixed pressure of the primary loop and not the complete primary loop flow path. To simulate the obstructed flow fault, the building operators installed a valve in series with the heat exchanger, as indicated. In the no-fault condition, with both primary and secondary pumps running, flow in the loop was approximately 1.7 L/s (27 gpm). With the fault in place, flow was reduced to 0.7 L/s (11 gpm), which is approximately 40% of nominal flow.

The experimental procedure was to introduce the simulated fault, turn CHWP-B on and off a few times, remove the fault, and again cycle CHWP-B. The resulting startup transients were recorded and then analyzed off-line.

Transducers were installed to measure current on two of the phases feeding the balanced, three-phase 480 V pump motor. Data were sampled with 12-bit resolution at 4000 Hz. A typical transient is shown in Figure 11. Inspection of the data revealed the presence of an unanticipated fault, consisting of an occasional line-cycle of zero current in the transient. This phenomenon was attributed to the contactor. Because contactor failure was not part of the model, transients with contactor problems were discarded. The remaining transients were labeled A, B, C, and D and will be referred to as such when results are presented. Transients A and B were collected under fault conditions.

A simple electrical model for the cooling system is an induction machine connected to an inertia with damping. In addition to the six parameters implied by this simplistic model, it is necessary to estimate the electrical angle when the motor was turned on and the initial speed of the motor. Extracting eight parameters from a simple transient as shown in Figure 11 is difficult.

It was assumed that the fault lay in the mechanical system (i.e., that the motor performance was the same when comparing two transients). A “joint,” two-transient model was formed

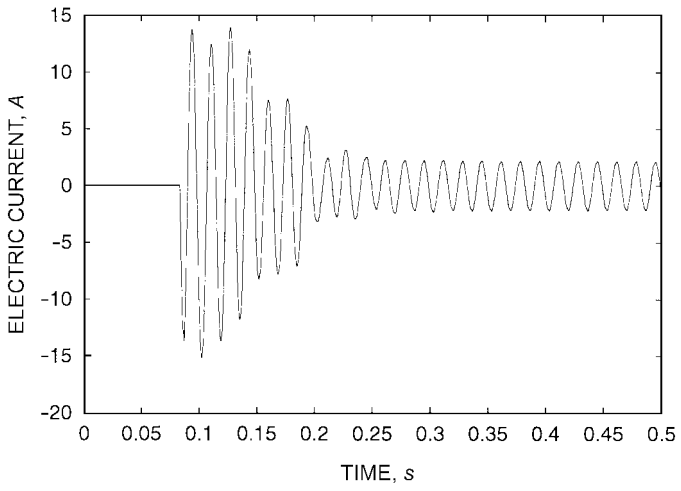


Figure 11. Typical Electrical Transient for CHWP-B Motor
The transient is the fault transient labeled "A" in Table 5.

where identical motor parameters were used for both transients, while the mechanical parameters were allowed to differ for each transient. Mathematically, the model consists of induction motor equations in what is known as *dq* space:

$$\frac{d}{dt}\begin{pmatrix}\lambda_{qs}\\ \lambda_{ds}\\ \lambda_{qr}\\ \lambda_{dr}\end{pmatrix} = \begin{pmatrix}v_{qs}\\ v_{ds}\\ 0\\ 0\end{pmatrix} - \begin{pmatrix}r_s i_{qs} + \omega\lambda_{ds}\\ r_s i_{ds} - \omega\lambda_{qs}\\ r_r i_{qr} + (\omega - \omega_r)\lambda_{dr}\\ r_r i_{dr} - (\omega - \omega_r)\lambda_{qr}\end{pmatrix} \tag{1}$$

where ω is the electrical frequency, ω_r is the rotor frequency, and the λs are the flux linkages (Krause 1986). The *dq* transform is a change of coordinates to a rotating frame, often applied to variables in electric machines to simplify analysis. The flux linkages are related to the currents by

$$\begin{aligned}\lambda_{qs} &= L_l i_{qs} + (L_l + L_m)(i_{qs} + i_{qr})\\ \lambda_{ds} &= L_l i_{ds} + (L_l + L_m)(i_{ds} + i_{dr})\\ \lambda_{qr} &= L_l i_{qr} + (L_l + L_m)(i_{qs} + i_{qr})\\ \lambda_{dr} &= L_l i_{dr} + (L_l + L_m)(i_{ds} + i_{dr})\end{aligned} \tag{2}$$

These electrical dynamics are influenced by a set of mechanical interactions, one for each transient:

$$\begin{aligned}\frac{d}{dt}\omega_1 &= 3K_1(\tau_1 - \beta_1\omega_1)\\ \frac{d}{dt}\omega_2 &= 3K_2(\tau_2 - \beta_2\omega_2)\end{aligned} \tag{3}$$

where the torque of electrical origin for the subscripted system is given by

$$\tau = \lambda_{qr} i_{dr} - \lambda_{dr} i_{qr} \tag{4}$$

The term β_1 (β_2) is a damping parameter and K_1 (K_2) is a parameter inversely proportional to the rotational inertia of the motor and pump. The output equations bring the simulation variables to the measured currents in the lab frame of reference. For each transient,

$$i = i_{ds}\cos[\omega(0)t + \phi] - i_{qs}\sin[\omega(0)t + \phi] \tag{5}$$

where ϕ is a parameter representing the electrical turn-on phase for each transient. The complete parameter vector of each joint model is then

$$\mu = [r_r r_s L_m L_l K_1 \beta_1 \omega(0)_1 \phi_1 K_2 \beta_2 \omega(0)_2 \phi_2]' \tag{6}$$

This parameter vector is then optimized to fit both transients in a given data set, as listed in Table 5. The result is that different mechanical parameters are obtained for each transient, while maintaining common electrical parameters. This model structure encapsulates the assumption that the difference between the two transients lies in the mechanical system. Parameter identification with different combinations of transients provides a useful crosscheck of the consistency of the estimates.

The model fit the data well, as shown in Figure 12. The residuals were small, but not structureless, as revealed by such typical residual analysis tools as the zero-crossing test and the Kolmogorov-Smirnov test (Johansson 1993). The model is simple and does not include complicated modeling of the interaction between water and pump. The residuals indicate that a more complicated model might be feasible with the data available. On the other hand, a more complicated model might cause identification problems with different data sets.

Parameters for the six data sets are shown in Tables 6 and 7. The data sets are separated by type, as given in Table 5. The parameters in Tables 6 and 7 agree well. In particular, the individual transients produce remarkably consistent estimates independent of the transients that they are paired with when fitting a joint model.

It is not necessarily important that the residuals are not structureless. The key issue for diagnostic purposes is the robustness of the parameters under perturbations of the measurements (i.e., whether parameter values are interpreted as faulty or not depends on the disturbances). This issue can be addressed by making parameter distribution estimates using synthetic data sets. Synthetic data set estimates of the distributions of the mechanical parameters, under fault and no-fault conditions with data set P_0, appear in Figure 13, which strongly suggests that detection of the fault would be successful under the presumed disturbance.

Table 5. Data Set Organization for Pump Tests

Data Set	Transient 1	Transient 2	Type
P_0	A	C	Fault/No-fault
P_1	A	D	Fault/No-fault
P_2	B	C	Fault/No-fault
P_3	A	D	Fault/No-fault
P_4	A	B	Fault/Fault
P_5	C	D	No-fault/No-fault

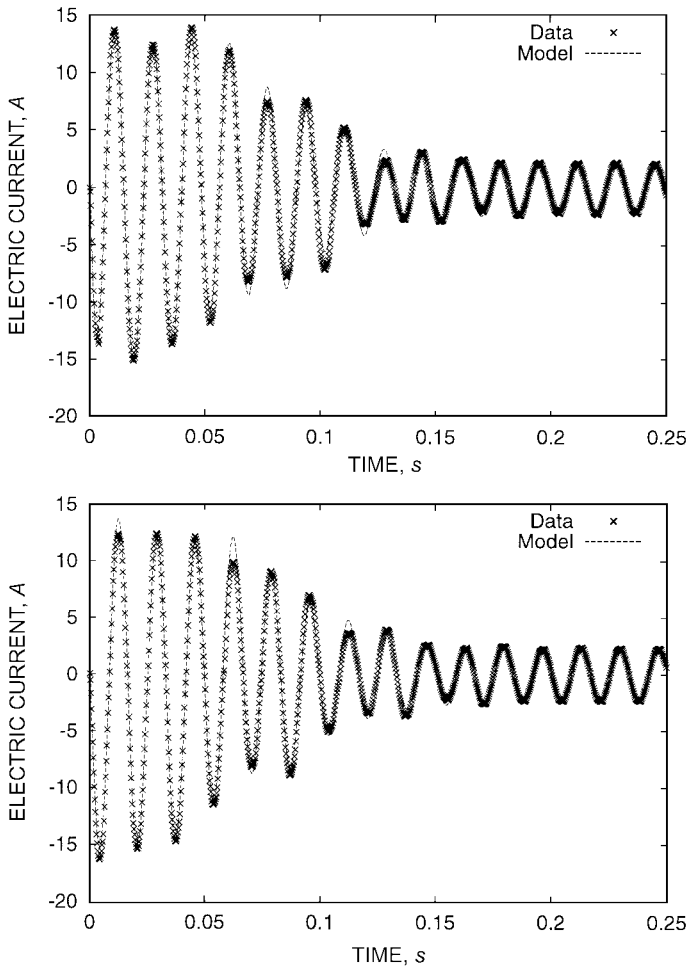


Figure 12. Comparison of Model to Experimental Data for Data Set P_0

Data with the reduced-flow fault are shown in the upper figure, while fault-free data are in the lower figure. The quality of fit is typical of the other data sets. While the fault-free and fault data appear visually similar, the differences are sufficient to produce significantly different values for the mechanical parameters.

Table 6. Electrical Parameter Estimates by Data Set for Pump Tests

	Transient		Electrical Parameters			
Data Set	**1**	**2**	r_r	r_s	L_m	L_l
P_0	A	C	1.548e + 01	1.424e + 01	7.266e – 01	3.046e – 02
P_1	A	D	1.487e + 01	1.418e + 01	7.050e – 01	3.096e – 02
P_2	B	C	1.529e + 01	1.421e + 01	7.205e – 01	3.061e – 02
P_3	A	D	1.511e + 01	1.418e + 01	7.132e – 01	3.077e – 02
P_4	A	B	1.545e + 01	1.380e + 01	7.324e – 01	3.072e – 02
P_5	C	D	1.499e + 01	1.462e + 01	7.048e – 01	3.046e – 02

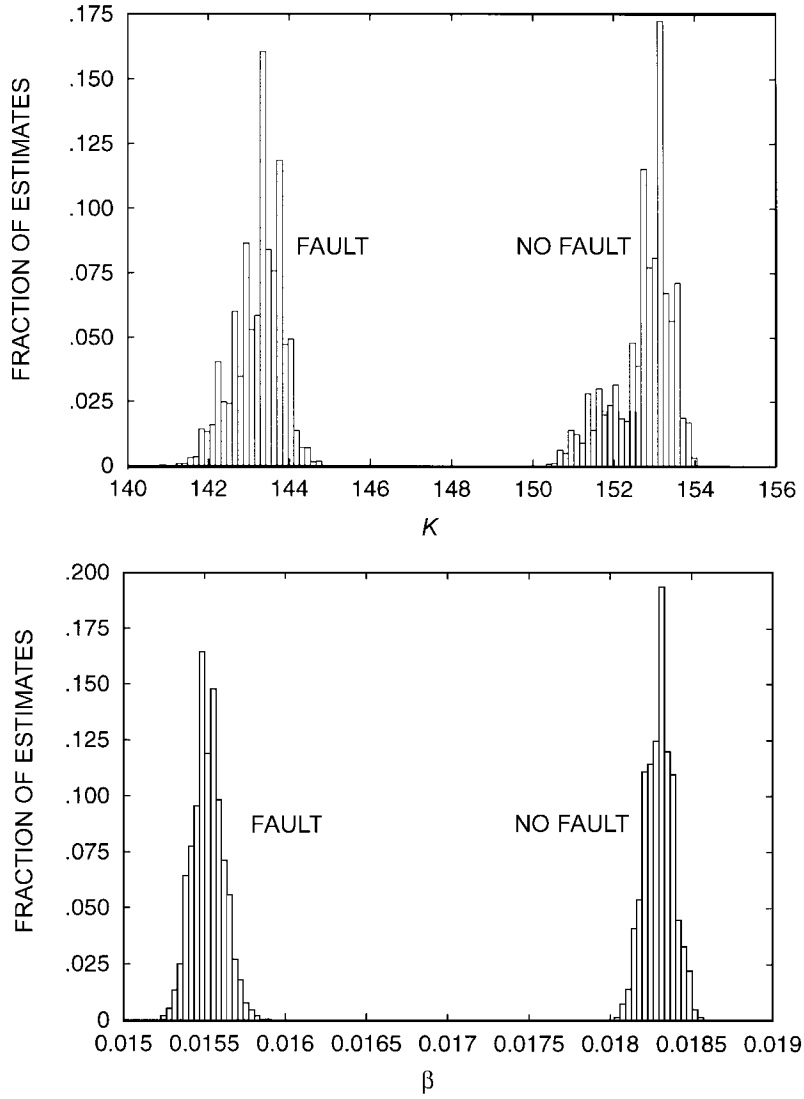


Figure 13. Distribution Estimates for Mechanical Parameters β and *K*

Table 7. Mechanical Parameter Estimates by Data Set for Pump Tests

	Transient		Mechanical Parameters			
Data Set	**1**	**2**	K_1	β_1	K_2	β_2
P_0	A	C	1.435e + 02	1.544e – 02	1.531e + 02	1.830e – 02
P_1	A	D	1.415e + 02	1.498e – 02	1.516e + 02	1.771e – 02
P_2	B	C	1.427e + 02	1.515e – 02	1.531e + 02	1.828e – 02
P_3	A	D	1.422e + 02	1.534e – 02	1.524e + 02	1.783e – 02
P_4	A	B	1.415e + 02	1.566e – 02	1.454e + 02	1.540e – 02
P_5	C	D	1.496e + 02	1.803e – 02	1.487e + 02	1.763e – 02

Fan Diagnostics

To investigate diagnostic possibilities in ventilation systems, data were collected from one of the test air-handling units. A diagram of the air-handling unit appears in Figure 14. Transients from the motor connected to supply fan B were measured, and two kinds of faults were considered. Unlike the pump situation, two functioning current transducers were available for measurement. For the first diagnostic, the outdoor air damper (indicated on the lower right of Figure 14) was opened to create a gross change in the flow characteristics of the system. The second diagnostic concern in the air-handling unit was the effect of a slipping fan belt, introduced in the test building by moving the motor to loosen the belt.

Developing an identification model of a complicated situation like Figure 14 is a challenging task. As a first assumption, controls (e.g., for the return fan) were assumed to operate on a time scale much slower than the induction motor startup transient. Also, other aspects of the system (e.g., state of doors and windows in rooms served by the air-handling unit) were assumed to remain constant during the test. The joint modeling technique used in the pump diagnostics was also used (i.e., the induction motor was assumed to be a constant for purposes of comparing two situations). The mechanical situation was modeled as in the pump scenario, i.e.,

$$\begin{aligned} \frac{d}{dt}\omega_1 &= 3K_1(\tau_1 - \beta_1\omega_1) \\ \frac{d}{dt}\omega_2 &= 3K_2(\tau_2 - \beta_2\omega_2) \end{aligned} \tag{7}$$

Unlike the pump diagnostics, two phases of current measurements were available for fan transients. These measurements were transformed to synchronously rotating *dq* space (Krause 1986) for identification.

Parameter estimates for the fan tests are given in Tables 8 and 9. As with the pump tests, transients were analyzed in pairs to detect changes in mechanical parameters associated with the presence of faults. Startup transient A refers to the no-fault situation, B corresponds to the open outdoor air damper (simulated by opening the mixing box door), C is the slipping belt, and D is both slipping belt and open door. The electrical parameters of Table 8 should be essentially the

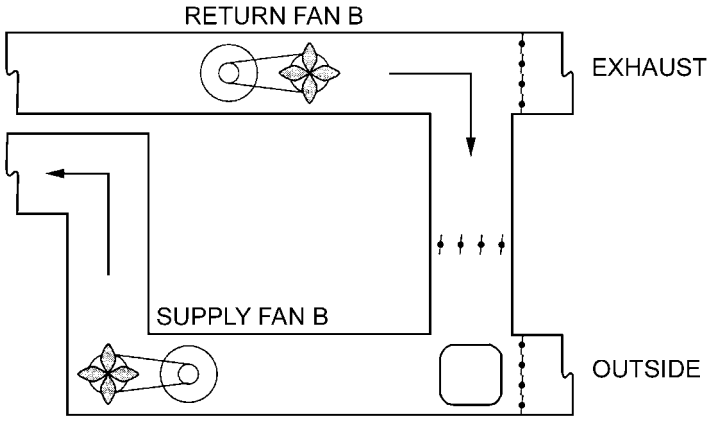


Figure 14. Schematic Diagram of AHU-B

Arrows indicate direction of airflow. Dampers on the right allow building air to be recirculated or mixed to varying degrees with outside air. The supply fan was used in the tests.

same for all combinations of data sets, because the same motor was used throughout. There is quite close agreement for parameters r_s and L_l and not such good agreement for the magnetizing inductance L_m and rotor resistance r_r. It is possible that the rotor resistance r_r might vary due to thermal effects, but the variation seen in the magnetizing inductance L_m is extreme. It seems likely that r_r and L_m are being influenced by the changes introduced in the mechanical part of the system. The most suspect combination in Table 8 is the first row.

The mechanical parameters for the six combinations of data sets are given in Table 9. In this table, parameters with subscript i correspond to the ith transient column. For example, β_2 in row three corresponds to an estimate for data set D (open door, loose belt) in conjunction with data set A (no-fault). Because the transients appear in different combinations in Table 9, a cross-validation check of sorts can be performed. Mechanical parameters corresponding to an individual transient (A, B, C or D) should be roughly the same. Figure 15 aids in this comparison by plotting the parameters corresponding to different mechanical situations. Note that the "no-fault" points in Figure 15 are quite distinct from the "fault" points. Also, the outlier in Figure 15 corresponds to the suspect row F_0 of Tables 8 and 9.

The parameters in Figure 15 and Table 9 make good physical sense. The situations where the fan belt is slipping show the lowest friction coefficient. As shown in Figure 14, the slipping fan belt situation is relatively close to the slipping fan belt and open door scenario. If the motor is not coupled to the fan, the duct configuration is irrelevant. Physical interpretation of K in Figure 15 is more involved. Because K is inversely proportional to the rotational inertia of the motor-fan system, larger values of K mean *less* inertia. The open door scenario has the most coupled inertia, followed by both slipping fan belt cases, and the no-fault situation has the least inertia. Because the slipping fan belt essentially uncouples the motor (at least for the startup transient), the motor should see only the inertia of the rotor and pulley. This hypothesis agrees well with

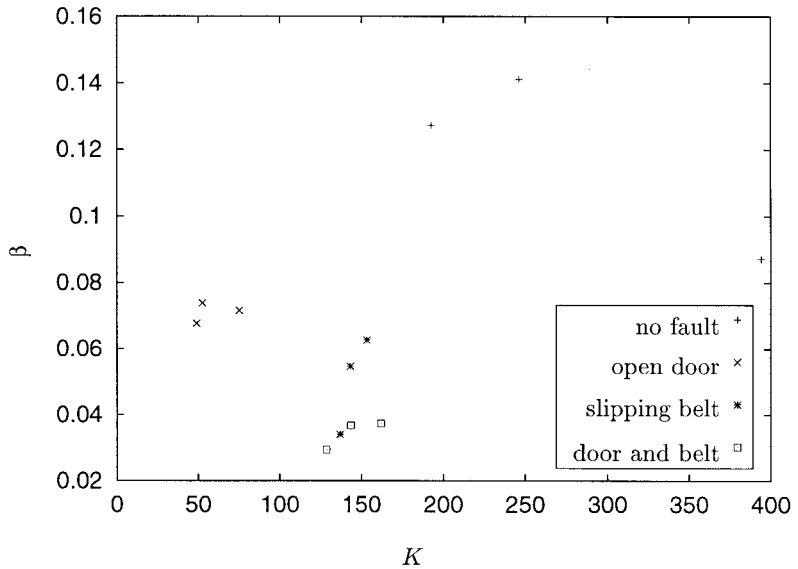


Figure 15. Scatter Plot Showing Parameters for Each Data Set When Estimated in Conjunction with Other Data Sets

The nine points in the lower left corner are fault points. Note that the rightmost point corresponds to the combination in the first row of Tables 8 and 9, which is suspect.

Table 8. Electrical Parameter Estimates for Fan Tests

	Transient		Electrical Parameters			
Data Set	**1**	**2**	r_r	r_s	L_m	L_l
F_0	A	B	2.35e – 01	8.07e – 01	5.84e – 02	4.62e – 03
F_1	A	C	3.75e – 01	8.15e – 01	9.72e – 02	5.13e – 03
F_2	A	D	3.99e – 01	7.95e – 01	9.14e – 02	4.99e – 03
F_3	B	D	2.89e – 01	7.98e – 01	1.17e – 01	4.97e – 03
F_4	C	B	3.02e – 01	8.06e – 01	9.79e – 02	4.58e – 03
F_5	C	D	3.26e – 01	8.03e – 01	1.52e – 01	4.78e – 03

Table 9. Mechanical Parameter Estimates for Fan Tests

	Transient		Mechanical Parameters			
Data Set	**1**	**2**	K_1	β_1	K_2	β_2
F_0	A	B	3.94e + 02	8.72e – 02	7.51e + 01	7.16e – 02
F_1	A	C	2.46e + 02	1.41e – 01	1.37e + 02	3.40e – 02
F_2	A	D	1.92e + 02	1.27e – 01	1.28e + 02	2.93e – 02
F_3	B	D	5.25e + 02	7.38e – 02	1.62e + 02	3.73e – 02
F_4	C	B	1.53e + 02	6.26e – 02	4.89e + 01	6.76e – 02
F_5	C	D	1.43e + 02	5.46e – 02	1.43e + 02	3.67e – 02

the close inertia values obtained for both slipping belt situations. In the no-fault scenario, the inertia values are *lower* than in the uncoupled case. This makes sense because the no-fault case is the only case where the return fan is supplying a force that tends to accelerate the rotor. The back pressure should not translate exactly into a change in inertia, but its gross effect may, particularly in the absence of a more complex model. When the outdoor air damper is open, the coupling of the supply fan to the return fan is greatly lessened, and the effective inertia includes the air column coupled to the fan.

The fit of the model to the observations was fairly good (see Figure 16). Note that these graphs show data that have been *dq* transformed (Krause 1986). Other mechanical models were attempted, including adding a $\gamma\omega^2$ term to Equation (5). These models did not offer much improvement in the fit, and in some cases it was clear that the data would not support a more complicated model.

CONCLUSION

Two methods have been developed for using electrical power measurements to detect a number of HVAC faults. The first, gray-box, method correlates steady-state power with such exogenous variables as flow, motor speed, or actuator control signal. Data that fall outside a tolerance band around a polynomial correlation indicate a fault. Faults can be diagnosed in limited cases where there are no other likely faults with the same power signature. The first method also detects and diagnoses cooling load faults via analysis of the cycling time of a reciprocating compressor. The second method fits high-speed power measurements taken during motor startup transients to a physical model that includes both electrical and mechanical parameters. Changes in the mechanical parameters indicate a fault. Both methods were demonstrated with data taken at a test building, where a modest number of artificial faults were introduced in the air handlers.

The gray-box FDD method was implemented with data from submeters and also, to a lesser extent, with data from centralized, high-speed electrical load monitors. Fault detection with this method has been limited to a relatively small range of possible faults in an air handler. Some

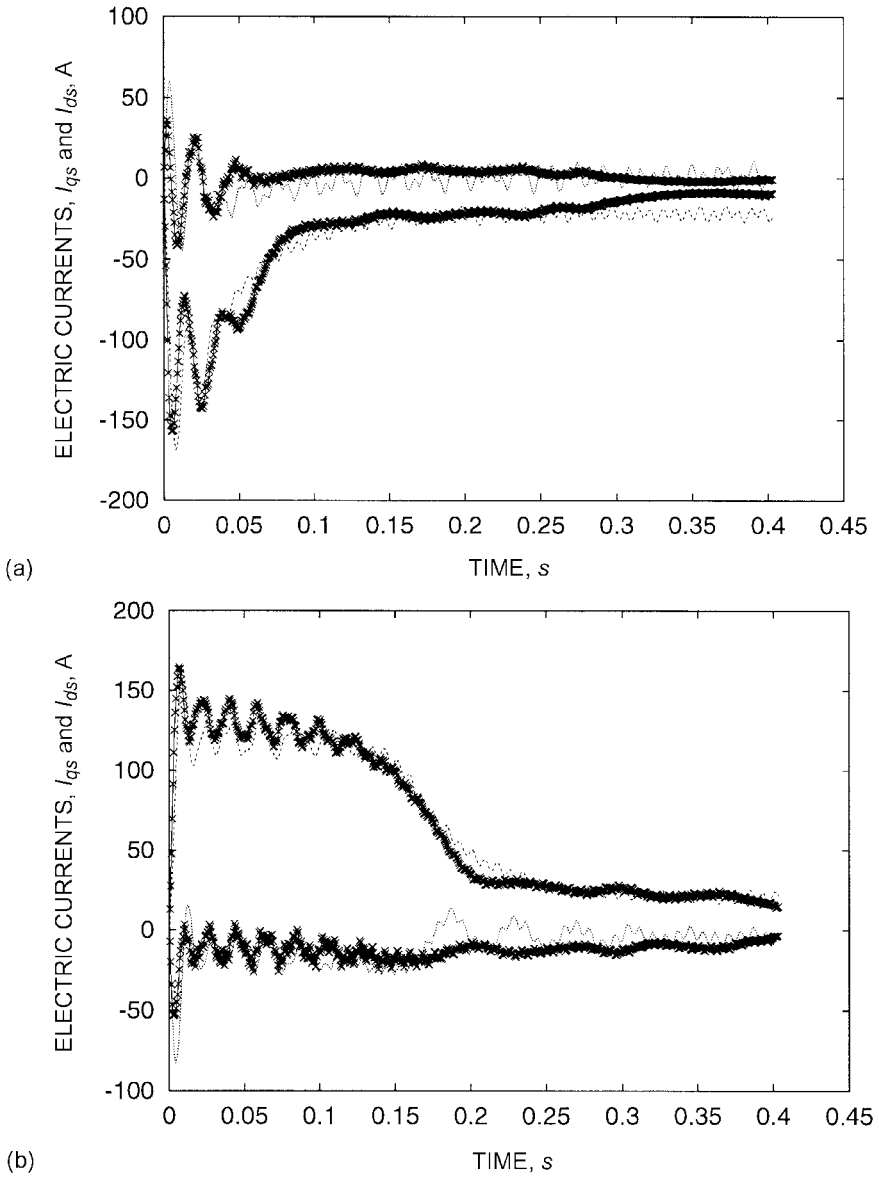


Figure 16. Fit of Joint Model to Experimental Data Sets
(a) Tight belt and 100% recirculation and (b) Tight belt and open outdoor air damper

forms of power correlations show more scatter than others and are less useful. For example, a correlation of fan power with flow has proved to be helpful in detecting changes in flow resistance, but it is more difficult to tightly correlate chiller power with outside temperature as a means of detecting a fault in an outdoor air temperature sensor. The strength of the correlation will have a strong impact on the extent to which the method can be extended to detect and diagnose a larger set of faults.

Detection and especially diagnosis capabilities were reduced when using a centralized, high-speed power meter. The centralized meter unavoidably has less resolution than dedicated sub-

meters, especially for relatively small pieces of equipment. Further, establishing the power correlations proved to be difficult. The central meter only collects data at equipment startup and shutdown. At the test building used in this study, startup signals for AHU fans were masked by the very slow control loop within the variable-speed drives. The secondary chilled water pump ran continuously, producing no power transients. Further, airflow, fan speed, and fan power at shutdown showed little deviation from day to day, due to similar system loads in late evening when the air handlers were turned off. Therefore, the central meter was used only to compare estimated fan power against a value considered to be normal, on the basis of shutdown measurements over a training period. Using only the power measurement made for an efficient means of fault detection but diagnosing the cause of the power deviation was not possible. For example, a deviation in fan power at the normal end-of-day shutdown could be due to a stuck damper or pressure sensor error. An alternative approach, now under investigation, requires that the fans be shut down during the day under different loads. With power data from the central meter and corresponding flow or speed data, power correlations can be established in the same manner as was done with submetered data.

While the use of electricity data for FDD was initially motivated by reduction in metering costs associated with the central meter, the first FDD method may be more appropriately implemented with submetered data if fault diagnosis and not just fault detection is highly valued. Less expensive sensors and sensor-communication systems (e.g., wireless sensors, to avoid wiring costs) will facilitate the use of submetered data.

The second method relies solely on submetered data and a physical model, which is more difficult to formulate than a simple power correlation but is also more powerful. It needs no correlative data, such as flow measurements, and can detect a fault with data collected over the very short period of time required for a motor to reach a steady speed. However, faults can only be detected at the time of motor startup, which introduces an unwanted lag between fault introduction and fault detection. Extending the method to work with data from a centrally located electrical meter, as will be attempted in the future, is possible in principle if the startup transient for a single motor is not masked by electrical noise introduced by other components. Limited laboratory tests have shown that a startup transient can be properly associated with a given motor even when another motor's startup partially overlaps the transient under investigation. However, such tests have not accounted for the sheer volume of events in a real building. Further, variable-speed drives are potentially a major stumbling block, because their prolonged startup cycle further masks the dynamics of the motor and driven load. This method is well suited to submetered power data and may appeal in the future as an onboard diagnostic method for equipment provided with a low-cost current transducer.

Both methods are potentially more robust than FDD methods that rely on temperature and flow sensors, in the sense that they do not require estimating small temperature differences with sensors that are subject to errors. Power measurements also avoid concerns about placement of flow and temperature sensors in large ducts or pipes. Both methods require testing in more buildings.

ACKNOWLEDGMENT

The authors warmly acknowledge the financial support of Honeywell, Inc., the California Energy Commission (via subcontracts from Lawrence Berkeley National Laboratory and Architectural Energy Corporation), and ASHRAE; the vision and guidance of Phil Haves, John House, and Kristin Heinemeier; and the expertise and dedication of Curt Klaassen and Andy Suby at the Iowa Energy Center's Energy Resource Station, the test building used in our field work.

NOMENCLATURE

T_{out}	outside dry-bulb temperature
$T_{balance\ point}$	balance point temperature, the outside dry-bulb temperature at which a building requires neither heating nor cooling
$T_{supply\ air}$	temperature of the supply air
$\Delta T_{supply\ fan}$	temperature rise across the supply fan
$T_{return\ air}$	temperature of the return air
λ	flux linkage for an induction motor
ω	frequency of excitation at the stator of an induction motor
ω_r	frequency of the rotor of an induction motor
i_{dr}	direct rotor current in the *dq* frame
i_{ds}	direct stator current in the *dq* frame
i_{qr}	quadrature rotor current in the *dq* frame
i_{qs}	quadrature stator current in the *dq* frame
L_i	leakage inductance of an induction motor
L_m	magnetizing inductance of an induction motor
r_r	electrical resistance of the rotor of an induction motor
r_s	electrical resistance of the stator of an induction motor
v_{ds}	direct stator voltage in *dq* frame
v_{qs}	quadrature stator voltage in *dq* frame
β	damping parameter for a system consisting of a motor and pump or fan
K	parameter inversely proportional to the rotational inertia of the motor and pump or fan
ϕ	the electrical turn-on phase for a motor startup transient
NILM	non-intrusive electrical load monitor

REFERENCES

Abler, C., R. Lepard, S. Shaw, D. Luo, S. Leeb, and L. Norford. 1998. Instrumentation for High-Performance Nonintrusive Electrical Load Monitoring. ASME *J. Solar Energy Engineering* 120:224-229.

Ahn, B.-C., J.W. Mitchell, and I.B.D. McIntosh. 2001. Model-Based Fault Detection and Diagnostics for Cooling Towers. *ASHRAE Transactions* 107(1).

Braun, J.E., J.W. Mitchell, S.A. Klein, and W.A. Beckman. 1987. Performance and Control Characteristics of a Large Cooling System. *ASHRAE Transactions* 93(1).

Chen, B. and J.E. Braun. 2001. Simple Rule-Based Methods for Fault Detection and Diagnostics Applied to Packaged Air-Conditioners. *ASHRAE Transactions* 107(1).

Dexter, A.L. and M. Benouarets. 1996. A Generic Approach to Identifying Faults in HVAC Plan. *ASHRAE Transactions* 102(1):550-556.

Dexter, A.L. and D. Ngo. 2001. Fault Diagnosis in Air-Conditioning Systems: A Multi-Step Fuzzy Model-Based Approach. *International Journal of HVAC&R Research* 7(1):83-102.

Englander, S. and L. Norford. 1992. Variable Speed Drives: Improving Energy Consumption Modeling and Savings Analysis Techniques. *Proceedings of the 1992 ACEEE Summer Study on Energy Efficiency in Buildings,* Pacific Grove, California 3:61-78.

Fels, M.F. 1986. PRISM: An Introduction. *Energy and Buildings* 9(1,2):5-18.

Hart, G.W. 1992. Nonintrusive Appliance Load Monitoring. *Proceedings of the IEEE* 80(12):1870-1891.

Haves, P., T.I. Salsbury, and J.A. Wright. 1996a. Condition Monitoring in HVAC Subsystems Using First-Principles Models. *ASHRAE Transactions* 102(1):519-527.

Haves, P., L.K. Norford, M. DeSimone, and L. Mei. 1996b. *A Standard Simulation Testbed for the Evaluation of Control Algorithms and Strategies*. Final Report for ASHRAE RP-825.

Haves, P., L.K. Norford, and M. DeSimone. 1998. A Standard Simulation Testbed for the Evaluation of Control Algorithms and Strategies. *ASHRAE Transactions* 104(1).

Hill, R.O. 1995. Applied Change of Mean Detection Techniques for HVAC Fault Detection and Diagnosis and Power Monitoring. M.S. thesis, Department of Architecture, Massachusetts Institute of Technology, Cambridge, Massachusetts.

House, J.M., H. Vaezi-Nejad, and J.M. Whitcomb. 2001. An Expert Rule Set for Fault Detection in Air-Handling Units. *ASHRAE Transactions* 107(1).

Hyvarinen, J., ed. 1996. *IEA Annex 25 Final Report*. VTT, Espoo, Finland.

Johansson, R. 1993. *System Modeling and Identification*. Prentice-Hall, Englewood Cliffs, New Jersey.

Krause, P.C. 1986. *Analysis of Electrical Machinery*. McGraw-Hill, New York.

Lee, W.Y., C. Park, and G.E. Kelly. 1996a. Fault Detection of an Air-Handling Unit Using Residual and Parameter Identification Methods. *ASHRAE Transactions* 102(1):528-539.

Lee, W.Y., J.M. House, C. Park, and G.E. Kelly. 1996b. Fault Diagnosis of an Air-Handling Unit Using Artificial Neural Networks. *ASHRAE Transactions* 102(1):540-549.

Leeb, S.B. 1993. A Conjoint Pattern Recognition Approach to Nonintrusive Load Monitoring. Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Leeb, S.B., S.R. Shaw, and J.L. Kirtley, Jr. 1995. Transient Event Detection in Spectral Envelope Estimates for Nonintrusive Load Monitoring. *IEEE Transactions on Power Delivery*. 10(3):1200-1210, July (also Paper 95 WM 042-2-PWRD IEEE PES Winter Meeting, New York, NY, January, 1995).

Li, X., H. Vaezi-Nejad, and J.-C. Visier. 1996. Development of a Fault Diagnosis Method for Heating Systems Using Neural Networks. *ASHRAE Transactions* 102(1):607-614.

Lorenzetti, D. and L.K. Norford. 1992. Measured Energy Consumption of Variable-Air-Volume Fans Under Inlet Vane and Variable-Speed Drive Control. *ASHRAE Transactions*. 98(2):371-379.

Luo, D., L.K. Norford, S.R. Shaw, and S.B. Leeb. 2002. Monitoring HVAC Equipment Electrical Loads from a Centralized Location—Methods and Field Test Results. Accepted for publication in *ASHRAE Transactions* and to be published in 108(1).

Norford, L.K. and N. Mabey. 1992. Nonintrusive Electric Load Monitoring in Commercial Buildings. *Proceedings of the Eighth Symposium on Improving Building Systems in Hot and Humid Climates*, Dallas, Texas.

Norford, L.K. and R.D. Little. 1993. Fault Detection and Load Monitoring in Ventilation Systems. *ASHRAE Transactions* 99(1):590-602.

Norford, L.K. and S.B. Leeb. 1996. Nonintrusive Electrical Load Monitoring. *Energy and Buildings* 24:51-64.

Norford, L.K., J.A. Wright, R. Buswell, and D. Luo. 2000. Demonstration of Fault Detection and Diagnosis Methods in a Real Building (ASHRAE 1020-RP). ASHRAE 1020-RP *Final Report*. ASHRAE, Atlanta. Report is posted on the TC4.11 website under www.ashrae.org.

Norford, L.K., J.A. Wright, R. Buswell, D. Luo, S.R. Shaw, and S.B. Leeb. 2002. Demonstration of Fault Detection and Diagnosis Methods for Air-Handling Units (ASHRAE 1020-RP). *International Journal of HVAC&R Research* 8(1).

Peitsman, H.C. and V.E. Bakker. 1996. Application of Black-Box Models to HVAC Systems for Fault Detection. *ASHRAE Transactions* 102(1):628-640.

Ruch, D. and D.E. Claridge. 1992. A Four-Parameter Change-Point Model for Predicting Energy Consumption in Commercial Buildings. *ASME Journal of Solar Energy Engineering* 114 (May):77-83.

Ruch, D., L. Chen, J.S. Haberl, and D.E. Claridge. 1993. A Change-Point Principal Component Analysis (CP/PCA) Method for Predicting Energy Usage in Commercial Buildings: The PCA Model. *ASME Journal of Solar Energy Engineering* 115 (May):77-84.

Salsbury, T.I. 1996. Fault Detection and Diagnosis in HVAC Systems Using Analytical Models. Ph.D. thesis, Loughborough University, U.K.

Seem, J.E., J.M. House, and C.J. Klaassen. 1998. Volume Matching Control: Leave the Outdoor Air Damper Wide Open. *ASHRAE Journal* 40(2):58-60.

Stylianou, M. and D. Nikanpour. 1996. Performance Monitoring, Fault Detection, and Diagnosis of Reciprocating Chillers. *ASHRAE Transactions* 102(1):615-627.

Tong, D. 1989. Intelligent and Healthy Buildings. *Proceedings of Designing for Environmental Quality 89*, Birmingham Polytechnic/RIBA, Solihull, U.K.

Tsutsui, H. and K. Kamimura. 1996. Chiller Condition Monitoring Using Topological Case-Based Modeling. *ASHRAE Transactions* 102(1):641-648.

Yoshida, H., T. Iwami, H. Yuzawa, and M. Susuki. 1996. Typical Faults of Air-Conditioning Systems and Fault Detection by ARX Model and Extended Kalman Filter. *ASHRAE Transactions* 102(1):557-564.

California Energy Commission
Public Interest Energy Research Program

HPCBS # E5P22T3c

HPCBS

High Performance Commercial Building Systems

Demonstration of Fault Detection and Diagnosis Methods for Air-Handling Units (ASHRAE 1020-RP)

Element 5 - Integrated Commissioning and Diagnostics
Project 2.2 - Monitoring and Commissioning of Existing Buildings

L.K. Norford, Massachusetts Institute of Technology
J.A. Wright and R.A. Buswell, Loughborough University, UK
D. Luo, United Technologies Corp.
C.J. Klaassen, IEC Energy Resource Station
A. Suby, Iowa State University

VOL. 8, NO. 1
HVAC&R RESEARCH
January, 2002

pier
BERKELEY LAB
United States Department of Energy
Science, Security and Energy: Powering the 21st Century

Acknowledgement

This work was supported by the California Energy Commission, Public Interest Energy Research Program, under Contract No. 400-99-012 and by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

VOL. 8, NO. 1 HVAC&R RESEARCH JANUARY 2002

Demonstration of Fault Detection and Diagnosis Methods for Air-Handling Units (ASHRAE 1020-RP)

L.K. Norford, Ph.D. Member ASHRAE
J.A. Wright, Ph.D., C.Eng Member ASHRAE
R.A. Buswell
D. Luo, Ph.D. Student Member ASHRAE
C.J. Klaassen Member ASHRAE
A. Suby Member ASHRAE

Results are presented from controlled field tests of two methods for detecting and diagnosing faults in HVAC equipment. The tests were conducted in a unique research building that featured two air-handling units serving matched sets of unoccupied rooms with adjustable internal loads. Tests were also conducted in the same building on a third air handler serving areas used for instruction and by building staff. One of the two fault detection and diagnosis (FDD) methods used first-principles-based models of system components. The data used by this approach were obtained from sensors typically installed for control purposes. The second method was based on semiempirical correlations of submetered electrical power with flow rates or process control signals.

Faults were introduced into the air-mixing, filter-coil, and fan sections of each of the three air-handling units. In the matched air-handling units, faults were implemented over three blind test periods (summer, winter, and spring operating conditions). In each test period, the precise timing of the implementation of the fault conditions was unknown to the researchers. The faults were, however, selected from an agreed set of conditions and magnitudes, established for each season. This was necessary to ensure that at least some magnitudes of the faults could be detected by the FDD methods during the limited test period. Six faults were used for a single summer test period involving the third air-handling unit. These fault conditions were completely unknown to the researchers and the test period was truly blind.

The two FDD methods were evaluated on the basis of their sensitivity, robustness, the number of sensors required, and ease of implementation. Both methods detected nearly all of the faults in the two matched air-handling units but fewer of the unknown faults in the third air-handling unit. Fault diagnosis was more difficult than detection. The first-principles-based method misdiagnosed several faults. The electrical power correlation method demonstrated greater success in diagnosis, although the limited number of faults addressed in the tests contributed to this success. The first-principles-based models require a larger number of sensors than the electrical power correlation models, although the latter method requires power meters that are not typically installed. The first-principles-based models require training data for each subsystem model to tune the respective parameters so that the model predictions more precisely represent the target system. This is obtained by an open-loop test procedure. The electrical power correlation method uses polynomial models generated from data collected from "normal" system operation, under closed-loop control.

Leslie K. Norford is an associate professor at the Massachusetts Institute of Technology, Cambridge, MA. **Jonathan A. Wright** is a senior lecturer and **Richard A. Buswell** is a research associate with Loughborough University, Department of Civil and Building Engineering, Loughborough, U.K. **Dong Luo** is a senior engineer with United Technologies Corporation. **Curtis J. Klaassen** is manager of Iowa Energy Center's Energy Resource Station. **Andy Suby** is project engineer at the Center for Sustainable Environmental Technologies, Iowa State University.

Both methods were found to require further work in three principal areas: to reduce the number of parameters to be identified; to assess the impact of less expensive or fewer sensors; and to further automate their implementation. The first-principles-based models also require further work to improve the robustness of predictions.

INTRODUCTION

In the last decade, a considerable amount of research has been carried out in the field of fault detection and diagnosis in HVAC systems. Hyvarinen and Karki (1996) summarized the efforts of an international collaboration [International Energy Agency (IEA) Annex 25] that listed typical faults in heating systems ranging from oil burners to district heating distribution systems; vapor-compression and absorption refrigeration machines; variable air volume (VAV) air-handling units (AHUs); and thermal storage systems. This work also produced a number of fault detection and diagnosis (FDD) methods:

Innovation approaches
- Physical models
- Time-series models
- State-estimation methods

Parameter-estimation approaches
- Methods based on physical models
- Characteristic curves
- Characteristic parameters

Classification approaches

Topological case-based modeling
- Artificial neural networks

Expert-system approaches
- Rule-based methods
- Associative networks

Qualitative approaches
- Formal qualitative approaches
- Fuzzy models

Many participants in this effort described their own methods; those of a generic nature or that focused on AHUs include Dexter and Benouarets (1996), Haves et al. (1996), Lee et al. (1996a, 1996b), Salsbury (1996), Yoshida et al. (1996), and Peitsman and Bakker (1996). These methods were developed and tested with simulations and laboratory test rigs where a high degree of experimental control can be applied. Such issues as interfaces with commercially available control systems, identification of the intended users and their needs, and methods for testing and evaluating the performance of FDD systems in systems installed in real buildings were not addressed. IEA Annex 34 followed Annex 25 and focused on the practical application of FDD techniques in real buildings (Dexter and Pakanen 2001).

ASHRAE sponsored the research described in this paper as a contribution to the global effort to demonstrate FDD methods in real buildings. This research focused on demonstrating FDD methods applied to systems installed in real buildings, encompassing the FDD methods, results from the trials, and the evaluation of the FDD method performance. The three objectives of the research were to

1. Demonstrate the operation of FDD methods for HVAC systems in a realistic building environment

2. Compare the performance of different FDD methods for different types of faults in AHUs and to assess the costs of their implementation
3. Archive and document the test data so they can be used to develop and test other FDD methods

Each of these objectives carried equal weight and each was largely met. This paper reports the completed work with respect to the first two objectives. A more substantial account and archived data are available in the final report to ASHRAE (Norford et al. 2000).

This research considered VAV AHUs only, although the methods can be applied to other types of HVAC systems. Both methods were based on a reference-model approach, where measurements from the system are compared to model predictions. A significant difference between the model predictions and the observations indicates that the system has deviated from the expected operating condition, which is taken to indicate the presence of a fault. The first-principles-based approach modeled the subsystem components and used fluid (air/water) quantity and property measurements and control-signal observations. The detection of faults focused on the effect of the fault condition on the output of the system process. The electrical power correlation method related electrical power measurements to fluid quantity and property measurements and control-signal observations; changes in the correlations were considered to be faults.

A description is given of the test building and the HVAC systems. The faults, their implementation and the FDD methods are described. The results from the four test periods are presented. The methods are evaluated on the basis of the accuracy, calibration, and cost of the required sensors; the ease of implementation of the methods, including selection and estimation of the model parameters and thresholds; the sensitivity and robustness of the methods in fault detection; and their success in fault diagnosis.

BUILDING, SYSTEMS, AND FAULTS

The fault-test program was conducted in a unique building that combined laboratory testing capability with real building characteristics and was capable of simultaneously testing two full-scale, commercial building systems side-by-side with identical thermal loads. The building was equipped with three VAV AHUs. Two were nominally identical (AHU-A and AHU-B), each serving four test rooms (Figure 1). The building had a true north-south solar alignment so that the pairs of test rooms had nearly identical exposures to the external thermal loads. The test rooms were unoccupied but were equipped with two-stage electric baseboard heaters to simulate thermal loads and with two-stage room lighting, both scheduled to represent various usage patterns. The third AHU (AHU-1) served the general areas of the facility including offices, reception space, a classroom, a computer center, a display room, service spaces, and a media center. A second classroom (not shown in Figure 1) was added to the east side of the building during the later stages of this project. AHU-1 was subject to variable occupant, lighting, external, and internal loads.

The test rooms, heating and cooling loops, and AHUs were well instrumented, including watt transducers for all components of interest. The A and B test rooms were individually controlled by a single commercial energy-management and control system (EMCS) and the general areas were controlled by a second EMCS. The building had a structural steel frame with internally insulated, pre-cast concrete panels, a flat roof, slab-on-grade flooring, and a floor area of 862 m^2 (9272 ft^2), including the new classroom. The east, south, and west test rooms each had 6.9 m^2 (74 ft^2) windows with double-layer, clear glass.

The heating plant consisted of a gas-fired boiler, circulation pumps, and the necessary control valves. Heating operation of the HVAC systems was not required as part of the tests conducted in this research, other than the preheating of the outside air during winter operation to simulate

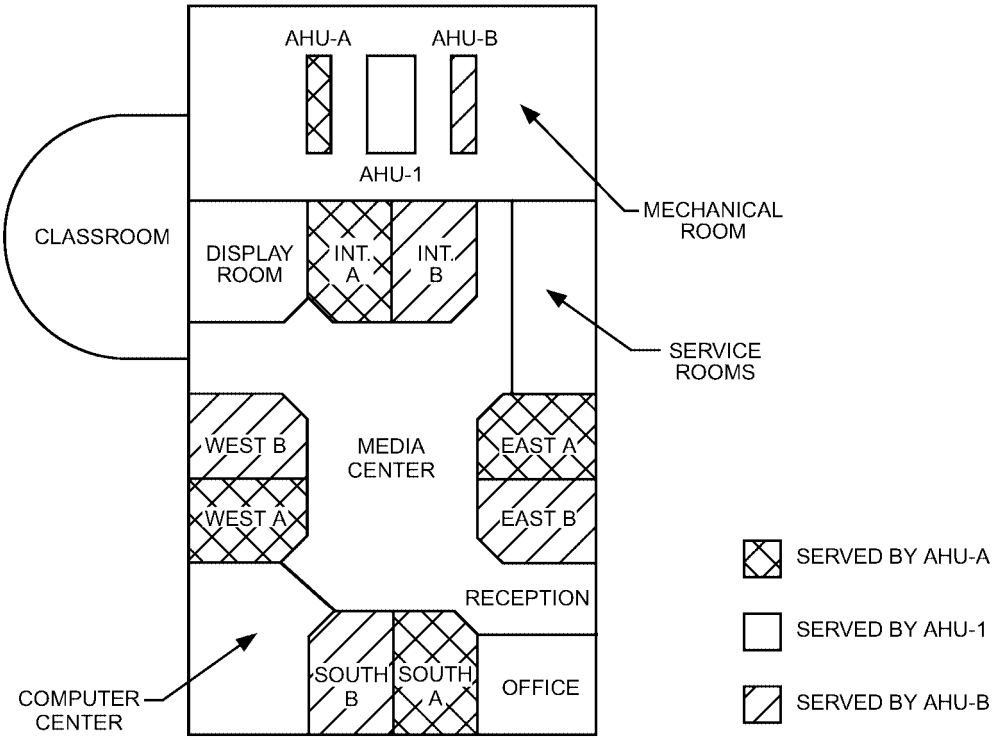


Figure 1. Plan of Test Building

higher outside temperatures and force the HVAC systems into economizer mode. The cooling plant (Figure 2), consisted of a nominal 35 $kW_{thermal}$ (10 ton) two-stage, reciprocating air-cooled chiller; a 525 $kWh_{thermal}$ (149 ton·h) thermal energy storage (TES) unit that was isolated from the cooling system for this research; and chilled water supplied by a central facility, with pumps, valves, and piping to circulate chilled water through the HVAC components.

The major components of the AHUs were the recirculated air, exhaust air, and outdoor air dampers; cooling and heating coils with control valves; and the supply and return fans (Figure 3). Ducts transferred the air to and from the conditioned spaces. Both the supply and return fans were controlled with variable-frequency drives. An additional heating coil was installed for this research on AHU-A and B, between the outside air (OA) inlet and the flow and temperature sensors. This coil was employed during the winter test period to preheat the outside air so as to force the control system into the free-cooling mode. AHU-A and B were identical, while AHU-1 was similar but larger to accommodate higher thermal loads. Air from the AHUs was supplied to VAV box units, each having electric or hydronic reheat.

The supply fan speed for all three AHUs was controlled to regulate supply duct static pressure. The AHU-A and AHU-B return fans were controlled to maintain a constant percentage of supply airflow; in AHU-1, the return fan control signal was a constant percentage of the supply fan control signal. The chilled water flow rate through the cooling coils in AHU-A and AHU-B was controlled by three-port mixing valves in a diverting application. A two-port valve was used in AHU-1.

The requirements for this project stipulated that a minimum of six faults be investigated, with at least two degradation faults and at least one fault in each of the three AHU subsystems (air-

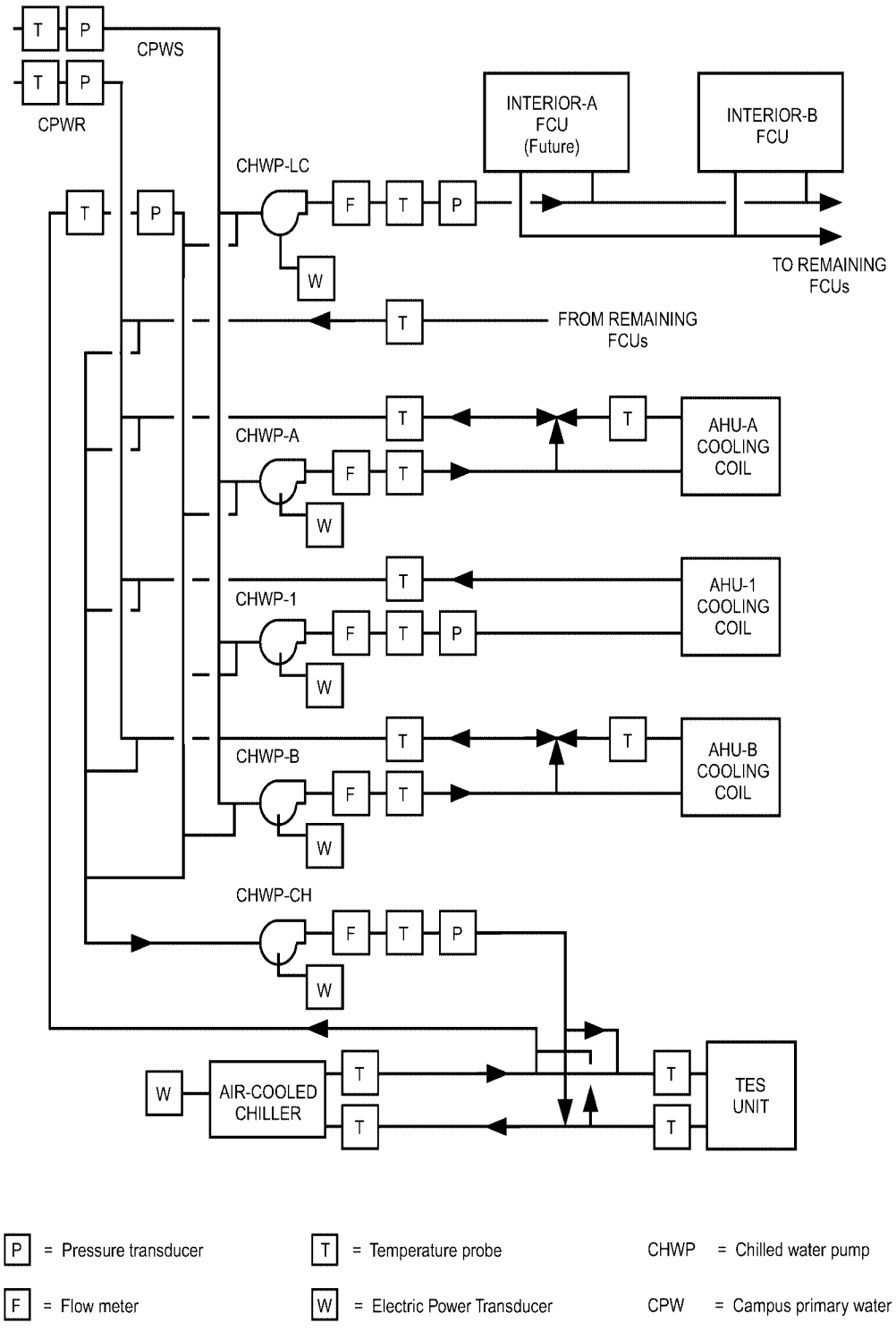


Figure 2. Chilled Water Flow Circuit in Test Building

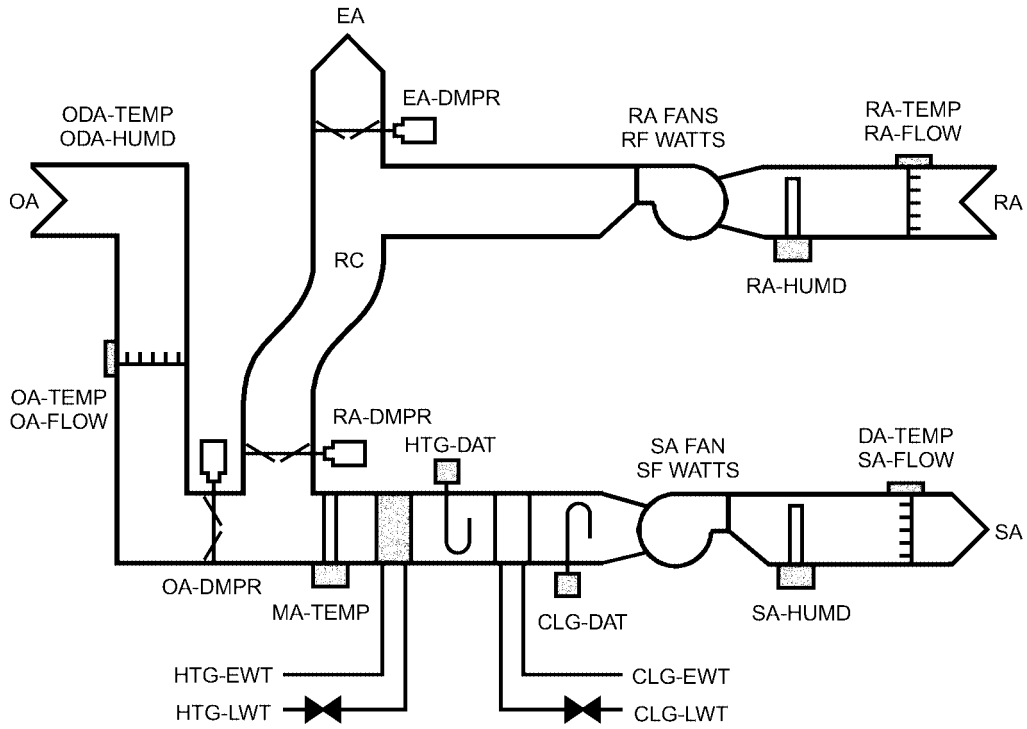


Figure 3. Air-Handling Unit in Test Building

mixing, filter-coil, and fan). Table 1 shows the seven selected faults and their method of implementation for AHU-A and AHU-B. Although faults implemented through software or easily disconnected hardware (such as actuator linkages) were readily introduced, others required substantial system modification, including the installation of bypass piping and additional valves.

It was necessary, in the context of the research, to ensure that both FDD methods were (in principle) capable of detecting all of the faults. Clearly, little would be learned from a series of null results. This criterion eliminated some faults, such as temperature sensor faults that the electrical power method would have difficulty detecting. This criterion was relaxed for tests with AHU-1, discussed later.

Table 2 indicates that each fault was implemented in at least two of the three test periods held during summer, winter, and spring seasons. Each test period consisted of a week of controlled tests, when the research staff of the test building introduced faults known to the investigators, a short analysis period, and a week of blind tests. For each blind test period, the list of possible faults was made known, but not the order of implementation or whether they were implemented at all. The lists for each season excluded faults that would not be seen in that season. For example, the recirculation damper would normally be fully open in hot weather (minimum outside air) and a damper leak could not be detected. Abrupt faults were typically implemented over a 24 h period while most degradation faults required three days, one for each of three stages. The three stages of the drifting pressure sensor fault were introduced over a single day.

Fault magnitudes were established during an initial period when the FDD methods were commissioned and the procedures for introducing faults and the HVAC systems were developed. The magnitudes of the degradation faults were selected such that it was anticipated that the two FDD methods *would* be able to detect the largest level, *should* be able to detect the middle level

Table 1. Method of Implementation of Faults

Fault	Type	Implementation
Air-Mixing Section		
Stuck-closed recirculation damper	Abrupt	Application of a control voltage from an independent source to maintain the damper in the closed position.
Leaking recirculation damper	Degradation	Removal of the recirculation damper seals, with one seal removed for the first fault stage, two for the second, and all seals for the third stage.
Filter-Coil Section		
Leaking cooling coil valve	Degradation	Manual opening of a coil bypass valve.
Reduced coil capacity (water-side)	Degradation	Manual throttling of the cooling coil balancing valve, to 70%, 42%, and 27% of the maximum coil flow of 1.7 L/s (27.5 gpm) for the three fault stages.
Fan		
Drifting pressure sensor	Degradation	Introduction of a controlled leak in the pneumatic signal tube from the supply duct static pressure sensor to the transducer, to a maximum reduction of 225 Pa (0.9 in. of water).
Unstable supply fan controller	Abrupt	Introduction of alternative gains for the PID controller that adjusts fan speed to regulate static pressure.
Slipping supply fan belt	Degradation	Adjustment of fan belt tension to reduce maximum fan speed by 15% at 100% control signal for the first stage and 20% for the second stage. The third stage had an extremely loose belt with variable fan speed.

and *could possibly* detect the lowest level. Fault magnitudes were consistent in each of the three test periods. Constant-magnitude faults provided a firmer basis for evaluating the FDD methods and were implemented with less difficulty than the variable-magnitude faults (a change of fault magnitude over different occurrences at different times) that would likely occur in practice. HVAC system commissioning consisted primarily of sensor calibration and establishing standard system operating configurations; the latter was required because the research facility altered the systems between test periods to meet the needs of other research programs. The configuration setup, which proved to be a major task for the test-building staff, encompassed fan control algorithms, isolation of the thermal storage tank (which provided a thermal capacitance that interfered with analysis of chiller cycling periods), and operating schedules for both HVAC equipment and false loads in the test rooms.

A more realistic set of blind tests was conducted with AHU-1, the air handler serving areas of the building occupied by research staff and classroom visitors. Four days of normal operation for training FDD methods and 17 days for fault introduction were included in a summer period of about six weeks. Building loads were not controlled and four of the six faults (listed in Table 3) had not been implemented in AHU-A and AHU-B and were completely unknown to the investigators. This test period was considerably longer than each of the three test periods on the matched AHUs, and increased the possibility of naturally occurring faults.

Two of the faults in this program produced signatures different from the naturally occurring faults they were intended to represent: the leaky cooling coil valve and the coil capacity fault. The leaking valve was implemented with a specially installed bypass valve that generated the same thermal effect as a leakage past a closed control port, but changed the flow resistance. The coil capacity fault mimicked the impact of water-side fouling on heat transfer across the cooling coil to some extent. A simpler alternative to replacing the existing coil with an older coil with tubes fouled with calcium carbonate was to close a valve in the inlet leg to the coil, thus increas-

Table 2. Faults Introduced into AHU-A and AHU-B During Three Blind Test Periods

Fault	Summer	Winter	Spring
Air-Mixing Section			
Stuck-closed recirculation damper	X	X	
Leaking recirculation damper		X	X
Filter-Coil Section			
Leaking cooling coil valve		X	X
Reduced coil capacity (water-side)	X		X
Fan			
Drifting pressure sensor	X	X	X
Unstable supply fan controller	X	X	X
Slipping supply fan belt	X	X	

Table 3. Faults Introduced into AHU-1 During Blind Test Period and Their Method of Implementation

Fault	Type	Implementation
Air-Mixing Section		
Stuck-closed recirculation damper	Abrupt	Application of a control voltage from an independent source to maintain the damper in the closed position for about 24 h
Stuck-open outside air damper	Abrupt	Application of a control voltage from an independent source to maintain the damper in the open position for 24 h
Filter-Coil Section		
Leaking heating coil valve	Abrupt	Adjustment of output voltage to the heating coil valve, causing it to unseat and leak for about 29 h
Fouled cooling coil	Degradation	Blockage of the cooling coil with a curtain drawn from the bottom to cover 25%, 50%, and 75% of the 61 cm (24 in.) coil in the three fault stages
Fan		
Drifting pressure sensor	Degradation	Introduction of a controlled leak in the pneumatic signal tube from the supply duct static pressure sensor to the transducer, with pressure reduced by 50, 100, and 150 Pa in the three fault stages (0.2, 0.4, and 0.6 in. of water) and each stage implemented for at least 6 h
Loss of control of supply fan	Abrupt	Supply fan VFD isolated from EMCS and operated at a constant speed for about 23 h

ing the resistance to water flow. This change in flow resistance became the basis for its detection with the electrical power FDD method.

Daily data sets for normal and faulty operation were assembled by the test-building staff from logs made by the EMCSs and were posted for electronic transfer to the investigators' home sites.

DETECTION AND DIAGNOSIS

The two fault detection methods compare the differences between the observed system behavior and a reference model of the system operation. The approaches differ significantly in how the fault effects are observed. The first-principles-based method considers the performance of the monitored system in terms of the system output useful to the air-conditioning process. In this case, the model predicts the temperature of the air or the static air pressure at the outlet of the component. A fault can be described in these terms as a degradation in the expected system performance.

The electrical power correlation method uses models derived from the system characteristics that relate electrical load to certain variables. This method predicts the expected power consumption. In this case a fault can be described as a change in the expected system energy consumption.

Both methods can take advantage of certain fault characteristics and not of others. The first-principles-based approach will always detect a degradation in the performance of the thermo-fluid system (as long as it is significant), regardless of the cause. The electrical power correlation method will not detect a fault that affects performance but has no effect on the electrical load. It is, however, predisposed to generating an operating cost associated with the faulty behavior and it can, in principle, detect faults associated with motors and drive trains that the first-principles-based approach cannot detect.

Fault diagnosis is also based on the information available to each method. The sensors required to implement FDD are one principal difference between the two methods. The first-principles-based approach uses measurements typically installed for control (temperatures, humidities, flow rates, etc.). One disadvantage with this approach is that in general terms there can be less control over the quality of these measurements in any given installation. The electrical power correlation method uses sensors over and above those normally installed, but these are more focused for the intended application and are not as susceptible to poor installation and maintenance.

First-Principles-Based Models with Thermo-Fluid Measurements

First-principles-based (or analytical) models can be used as a reference for the "correct" or expected operation of a HVAC system. The approach used in this research relied on the sensors typically installed in most VAV systems for control. Three subsystem models [described more fully in Norford et al. (2000)] were used to implement the FDD scheme: a fan/duct model of the air system, an economizer model, and a model of the coiling-coil subsystem. Figure 4 demonstrates the modeling arrangement. The black dots indicate where the comparisons to the observations from the real system were made, and hence where the fault detection for each subsystem was focused. Simple, steady-state simulations of the subsystems are formed by the models, which are based on the following principles:

- The fan/duct model is based on the fan laws and simple quadratic expressions for the change in system resistance and predicts the supply air static pressure.
- The economizer model is based on the analytical representation of the mixed-air condition as a function of damper position and the inlet temperatures and humidities. The model also includes an actuator model.
- The cooling coil model is based on the SHR method effectiveness-NTU heat and mass transfer calculation method [similar to the ASHRAE 3-line method; a review of both methods is given by Stephan (1994)]. The subsystem model also includes fan temperature rise and models for the control valve and the actuator.
- The fan-temperature-rise model is a simple addition to the air temperature, linearly dependent on the fan-control signal.
- The valve model is based on a first-principles analysis of the water-circuit resistance with respect to the control valve. The model predicts the mass flow rate of water through the coil, a typically unmeasured variable.
- The actuator model is an analytical representation of the movement of the actuator in response to a control signal. This models the dead-bands at either end of the operating range and any hysteresis (slack in the linkage) that may be present in the system.

The models have parameters for which values must be estimated for a specific system. The parameters are designed, as far as possible, to represent some tangible system characteristic and give greater precision in prediction. An example of this is the actuator "low activation point"

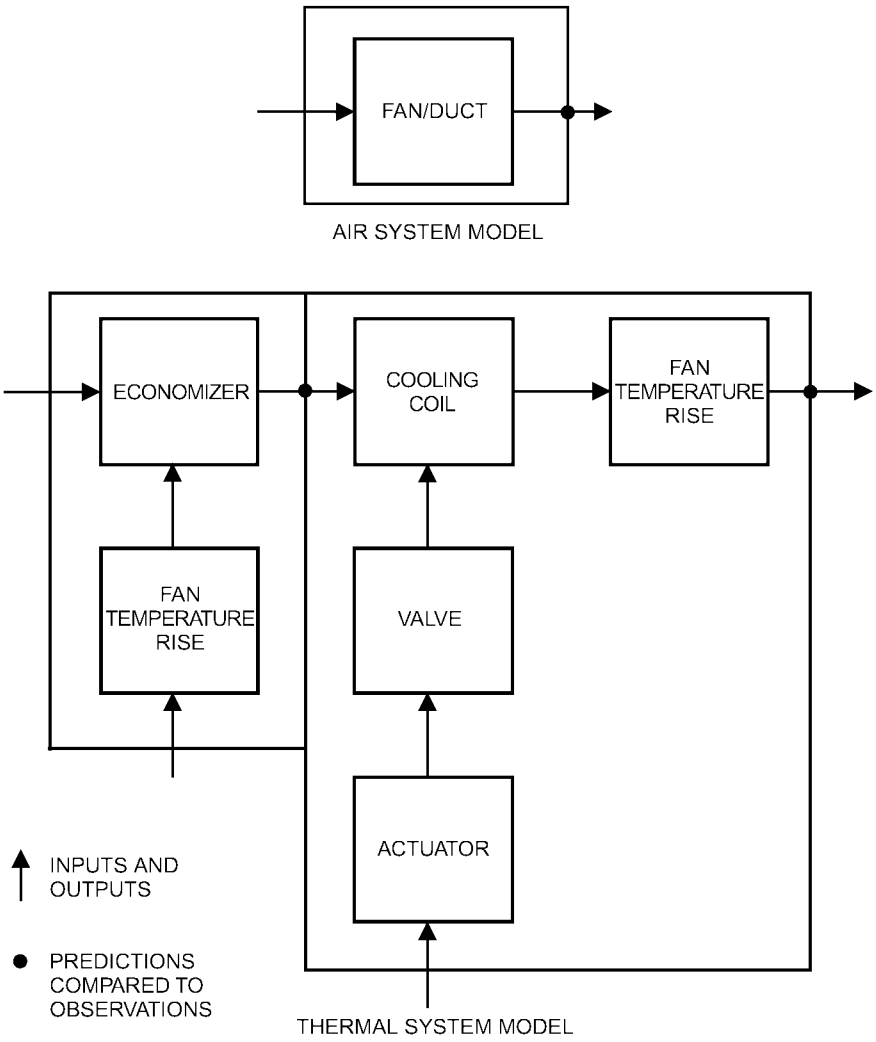


Figure 4. First-Principles-Based Model Functionality

parameter, which describes the value of the control signal required before the valve stem starts to move as it is opened from the closed position (i.e., dead-band). Some of the model parameters can be obtained from design information or inspection of the installed system. The face area and number of rows and circuits in the cooling coil are examples. The remaining parameters are identified simultaneously for each subsystem model. The model parameters are estimated by applying a nonlinear optimization technique, minimizing the model prediction errors in a least-squares sense. The data used for this procedure were generated by applying a sequence of open-loop "steps" in the inputs to capture the system characteristics when the system was in a normal (fault-free) condition.

With the system models characterized, model predictions can be applied to the observations to generate the "prediction error" (demonstrated in the top halves of Figures 5 and 6). The models only apply to observations that are close to steady state and a steady-state filter removes data containing transients. A lack of steady-state data was used to identify the presence of oscillatory, or unstable, control.

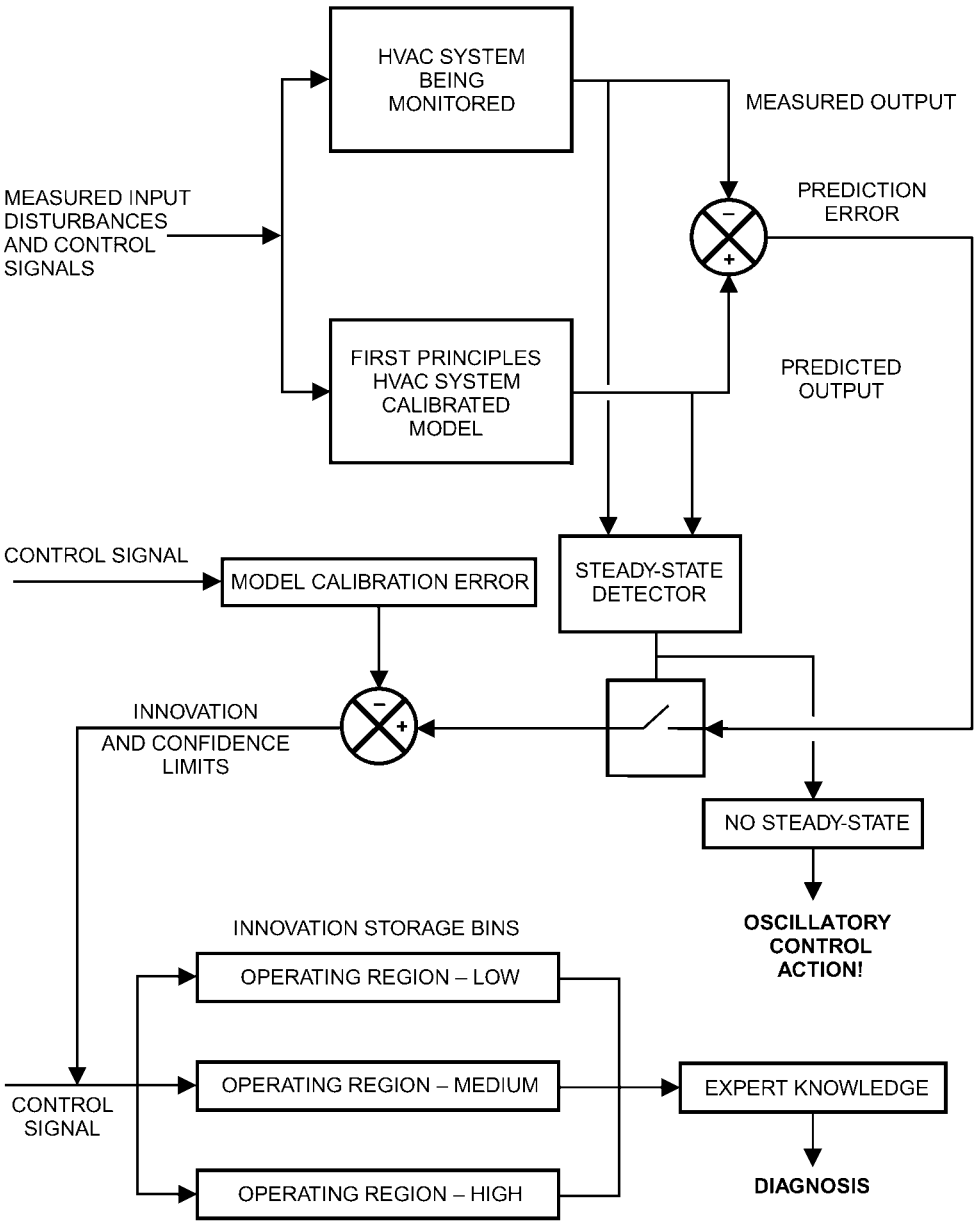


Figure 5. Method for Fault Detection and Fault Diagnosis Using Expert Rules

Some prediction error will always exist due to uncertainties in the measurements and unmodeled system disturbances. Statistically based thresholds are applied to the prediction error, such that a certain magnitude of error is required before triggering an alarm.

Once an alarm has been identified, the cause is diagnosed. Two methods were investigated in this work, fault diagnosis by (1) expert rules and by (2) recursive parameter estimation. The schemes are shown in Figures 5 and 6, respectively. Figure 5 shows that the "innovations" (the magnitude of the prediction error over and above the thresholds) were split into three "bins."

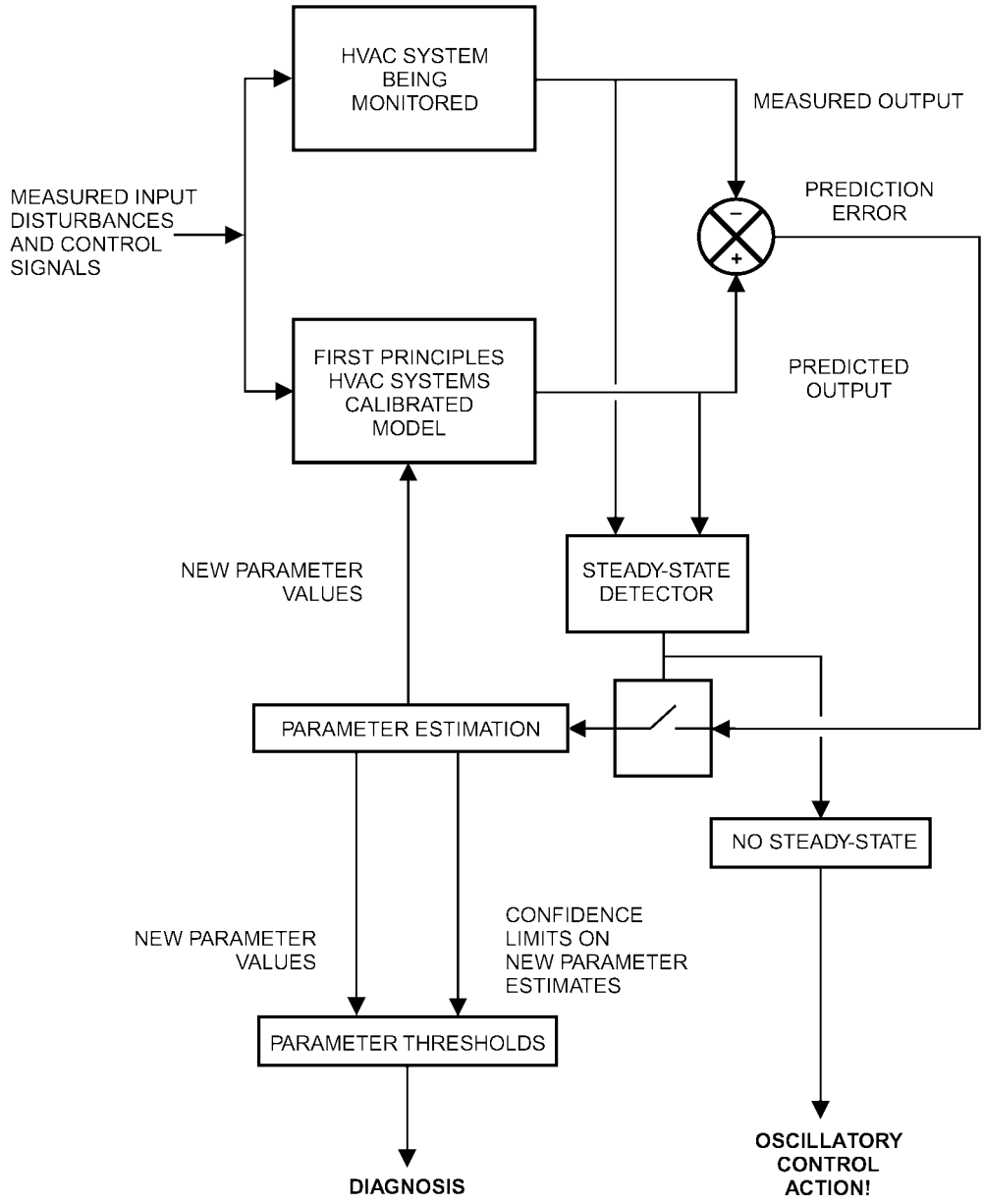


Figure 6. Method for Fault Diagnosis by Recursive Parameter Estimation

The bins contain the average magnitudes of the innovations, exponentially weighted with age. Each bin represents a portion of the operating space of the monitored process. Crisp expert rules were then applied to the average values in these bins to determine the cause of the fault.

In the recursive parameter estimation scheme, some of the parameters are designed to represent the effects of the faults on the system output. These parameters are recursively re-estimated to track the developing fault. The algorithm minimizes the prediction error and uses the sensitiv-

ity coefficients of the fault parameters with respect to the model output to drive the estimation procedure (Salsbury 1996). As the fault develops, the current fault parameter values implicitly describe the state of the system and hence, diagnose the state of the system.

Gray-Box Correlations with Electrical Measurements

The electrical power correlation FDD method produced prediction errors in electrical power. The method also made use of statistically derived confidence intervals for predictions of performance under normal operation. The method is a semiempirical approach that correlates measured fan or pump power with such exogenous variables as airflow, motor speed control signals and actuator position control signals. Power correlations were third-order polynomials; confidence intervals reflected the influence of disturbances during training periods, such as those due to normal variation in damper positions. Oscillatory power data, indicative of unstable local-control loops, were detected via a calculation of signal variance over a sliding window of data points; this calculation effectively acted as a steady-state filter by excluding oscillatory data from comparison with power correlations established during the training phase.

Analysis of chiller power (associated with the economizer and cooling coil valve leakage faults) was more difficult than for fans (air system faults) and pumps (cooling coil undercapacity fault), for two reasons. First, the chiller in the test building was a two-stage reciprocating unit with discrete power levels (0, 5, and 10 kW). In principle, it was possible to time-average the power to obtain a continuous power variable suitable for the same sort of power correlations used for fans and pumps. In practice, the cycling frequency was often long (i.e., on the order of 30 minutes), making it impossible to calculate a short-term power average needed for reasonable correlation with driving variables. Second, chiller power was strongly influenced by environmental conditions (expressed by dry-bulb and wet-bulb temperatures and solar radiation) and by building internal loads. These variables are not all easily measured. Even those that are directly and simply measurable require sensors that have a cost associated with them and are subject to errors. It is necessary to either include these influences in a model of chiller power or exclude them and limit the analysis of chiller power to narrow and known operating conditions.

The FDD method developed and applied in the test building assessed chiller cycling periods under two low-load conditions where it was, in principle, possible to discern a change in chiller loading due to damper and valve leaks.

The method relied heavily on one-minute-average data from installed power transducers to assess the benefits of such data and set the stage for a future cost-benefit analysis. A detailed discussion of the method is presented in a companion paper (Shaw et al. 2002), which includes examples of power correlations, detection of chiller cycling, and analysis of power oscillations.

Table 4 summarizes the types of electrical power analyses used in this method, along with a list of possible faults that each analysis can detect. The list of faults is not exhaustive but is long enough to indicate the difficulties in distinguishing a particular fault from other possible causes of the same deviation between predicted and measured electrical power.

Expert rules were used for limited fault diagnosis. Table 4 indicates how rules can distinguish a slipping fan belt from a fault caused by a change in flow resistance: the former leads to power measurements that differ from the predicted value for a given motor speed while the latter does not. (For the slipping fan belt, the reduction in power due to reduced airflow was a stronger effect than an increase in thermal dissipation from the fan belt itself, which became quite hot when slipping.) Although careful analysis of fan curves indicates that this statement is not entirely true, the impact of a change in duct pressure on the power-speed correlation is sufficiently minimal to be of no concern. As a second example, the leaky recirculation damper or cooling coil valve will affect chiller power but not fan power for a given airflow.

Table 4. Nonexhaustive Listing of Faults Associated with a Given Electrical Power Signature

Type of Electrical Power Analysis	Possible Faults Causing a Deviation Between Predicted and Measured Electrical Power
Polynomial correlation of supply fan power with supply airflow	Change in airflow resistance, possibly due to stuck air-handler dampers or fouling of heating or cooling coils
	Static pressure sensor error (affects portion of fan power due to static pressure)
	Flow sensor error
	Power transducer error
	Change in fan efficiency, caused by change in blade type or pitch, or use of VFD in lieu of inlet vanes
	Change in motor efficiency
Polynomial correlation of supply fan power with supply fan speed control signal	Slipping fan belt
	Disconnected control loop (fan speed differs from control signal)
	Power transducer error
	Change in fan efficiency
	Change in motor efficiency
Polynomial correlation of chilled water pump power with cooling coil control valve position control signal	Change in water flow resistance, possibly due to constricted cooling coil tubes or piping
	Disconnected control loop
	Power transducer error
	Change in pump efficiency
	Change in motor efficiency
Detection of change in cycling frequency for two-stage reciprocating chiller	Leaky cooling coil valve
	Leaky recirculation damper
Detection of power oscillations	Unstable local-loop controller

Ideally, a given power signature would be associated with nonoverlapping lists of faults, providing a high level of "orthogonality" useful for fault diagnosis. As can be seen in Table 4, this ideal was not achieved. Errors in power transducers and changes in fan or motor performance will affect both types of fan power correlations. In the blind tests, it was possible in some cases to distinguish faults associated with a given power correlation by limiting the analysis to a small range of the correlation or to a narrow band of another variable:

- The stuck-closed recirculation damper could be distinguished from the pressure sensor offset in the test building via analysis of power at low airflows. The impact of the stuck-closed recirculation damper fault was exacerbated in the early evening, when the air handlers were still running but, in cold or hot weather, the outdoor air damper was fully closed. The supply fan then drew air across two closed dampers. There is substantial variation across buildings in control strategies at the beginning and end of the working day and it is difficult to generalize such an approach. Even in the test building, this strategy could not be generalized to AHU-1, which operated continuously.
- The leaky cooling coil valve and the leaky recirculation damper could be distinguished with the help of measurements of the valve position control signal and outdoor temperature (Shaw et al. 2002).

This approach to fault diagnosis, while unable to distinguish a large set of possible faults, was easily implemented in rules. For the blind tests, where the number of possible faults was limited,

one rule was "if the electrical power exceeds the confidence interval and the airflow is less than a threshold, then the fault is a stuck-closed recirculation damper." More generally, the "if-then" statement could be modified to include a larger list of possible causes of the detected fault.

RESULTS

The results from the FDD trials on AHU-A/B and AHU-1 are shown in Tables 5 through 8. The original intent of the first-principles-based approach was to not use the mixed-air temperature sensor because this is not commonly available. The predictions of the mixed-air humidity ratio and temperature by the economizer model were used as inputs to the cooling coil model. The magnitudes of prediction errors associated with normal operation in the mixing box and cooling coil (due to the influence of unmodeled disturbances) led to a reduction in sensitivity to fault detection. There was also a significant reduction in the isolation of the cause of faults (faults possibly being in one of two subsystems), which led to ambiguous diagnoses. The model parameters were re-estimated using the mixed-air temperature measurement and the summer tests were rerun with the addition of this measurement. The analyses for the other season were carried out using the same models.

Table 5 describes the first blind test period, conducted in summer conditions on AHU-A and AHU-B. Tables 6 and 7 describe the winter and spring blind test periods. The tests on AHU-1 are summarized in Table 8; for these tests the first-principles-based FDD approach gave a single diagnosis through application of the RPE method and expert rules.

For both the first-principles-based models and the electrical power correlation method, the results are discussed first for AHU-A and AHU-B and then AHU-1. Where individual faults are highlighted, the order is the same as that presented in Tables 2 and 3.

First-Principles Models with Thermo-Fluid Measurements

In general, the first-principles-model-based FDD method proved to be effective in the detection of the faults implemented on AHU-A and B. All faults were detected in each season they were implemented, with the exception of the leaking recirculation damper and leaking cooling coil valve. Diagnosis was less reliable, in that no single method of diagnosis (expert rules or recursive parameter estimation) was able to provide a diagnosis for all fault conditions.

A leaking recirculation damper can be expected to produce small differences between the expected and observed mixed-air temperature. This is, however, a function of the temperature difference between the ambient and return air streams (when these temperatures are equal, no faults can be detected using temperature measurements) and the size of the leakage. In order to detect these changes, the model of the economizer must be quite precise in its prediction of the mixed-air temperature. The principal factors affecting the precision of the model were

- Stratification at the locations of the temperature sensors (return and mixed), which affects the calibration of the model parameters and subsequent calculations of prediction error.
- Localized offsets caused by differences between the measured ambient-air temperature and the temperature of the ambient air entering the inlet duct, some distance away. The effect was similar to the above point.
- Unmodeled effects, of which there were two primary sources:
    - The pressure/resistance characteristics for the fan/duct system were not constant as a result of the fan control strategy (return fan runs at a fixed percentage of the supply air volumetric flow rate). This meant that at different fan speeds, different proportions of the ambient air and return air were mixed for a given damper position. Prediction errors are generated because the model assumes that these proportions are unchanging (i.e., that the system is well balanced).

Table 5. Detection and Diagnosis of Faults During Summer Blind Test Period for AHU-A and AHU-B

Test Day	AHU A	AHU B	Physical Models: Detect from Innovations	Physical Models: Diagnose from Expert Rules	Physical Models: Diagnose from Recursive Parameter Estimation	Electrical Power Models: Detect	Electrical Power Models: Diagnose
1	Slipping fan belt (Stage 1)		No	—	—	Yes	Yes
		Reduced cooling coil capacity (Stage 1)	Yes	Yes	No	Yes	Yes
2	Slipping fan belt (Stage 2)		No	—	—	Yes	Yes
		Reduced cooling coil capacity (Stage 2)	Yes	Yes	No	Yes	Yes
3	Slipping fan belt (Stage 3)		Yes	No	Yes	Yes	Yes
		Reduced cooling coil capacity (Stage 3)	Yes	Yes	No	Yes	Yes
4	No fault		No fault	—	—	No fault	—
		Unstable pressure control	Yes*	—	—	Yes	Yes
5	Unstable pressure control		Yes*	—	—	Yes	Yes
		No fault	No fault	—	—	No fault	—
6	Static-pressure sensor offset (Stages 1-3)		Yes	No	Yes	Yes	Yes
		Stuck-closed recirculation damper	Yes	No	Yes	Yes	Yes
7	Stuck-closed recirculation damper		Yes	Yes	Yes	Yes	Yes
		Slipping fan belt (Stage 1)	No	—	—	No	—

*The unstable pressure controller was detected via the steady-state filter, which indicated that the measured pressure was in a dynamic state for a prolonged period. No detection or diagnosis method was applied, because the measurements did not pass the filter.

- At certain conditions, the fan/duct system imbalance also caused ambient air to flow in through the exhaust grille, resulting in a higher proportion of the ambient air in the mixed air than expected by the model. This problem was exacerbated in the system under investigation because of the increased resistance to airflow through the ambient air inlet duct due to the installation of the preheat coil.

These problems can be considered to be system faults and not problems with the FDD method. The last point highlights the disadvantages of an analytical modeling approach applied to observations of the result of the process, rather than modeling the process itself. Given a more detailed fan/duct system model, it may be possible to predict the airflow rates within the system (within a tolerable degree of uncertainty), which would eliminate the need for the mixing box model as presented in this work.

Table 6. Detection and Diagnosis of Faults During Winter Blind Test Period for AHU-A and AHU-B

Test Day	AHU A	AHU B	Physical Models: Detect from Innovations	Physical Models: Diagnose from Expert Rules	Physical Models: Diagnose from Recursive Parameter Estimation	Electrical Power Models: Detect	Electrical Power Models: Diagnose
1	No fault		No fault	—	—	No fault	—
		Stuck-closed recirculation damper	Yes[1]	No	Yes[2]	Yes	Yes
2	Leaking cooling coil valve (Stages 1-3)		No	—	—	Yes	Yes
		Slipping fan belt (Stage 1)	No	—	—	No	—
3	Leaking recirculation damper (Stage 1)		No	—	—	Yes	Yes
		Slipping fan belt (Stage 2)	No	—	—	No	—
4	Leaking recirculation damper (Stage 2)		No	—	—	Yes	Yes
		Slipping fan belt (Stage 3)	Yes	Yes[3]	Yes	Yes	Yes
5	Leaking recirculation damper (Stage 3)		No	—	—	Yes	Yes
		Static pressure sensor offset (Stages 1- 3)	Yes	Yes	Yes	Yes	Yes
6	Static pressure sensor offset (Stages 1- 3)		Yes	Yes[4]	Yes[5]	Yes	Yes
		Unstable pressure control	Yes	Yes	Yes	Yes	Yes
7	Unstable pressure control		Yes[6]	—	—	Yes	Yes
		No fault	No fault	—	—	No fault	—

[1]Alternative diagnosis of static pressure sensor drift.
[2]Alternative diagnosis of slipping fan belt.
[3]Alternative diagnosis of unknown mixing box fault.
[4]Alternative diagnoses of excessive control dynamics and excessive outside air (due to flow of outdoor air into the exhaust damper, an actual—not artificial—system fault).
[5]Alternative diagnosis of slipping fan belt.
[6]The unstable pressure controller was detected via the steady-state filter, which indicated that the measured pressure was in a dynamic state for a prolonged period.

Estimating the parameters for the economizer model means that either the system has to be balanced correctly or the training data (ideally) need to encompass the complete range of expected airflow/resistance characteristics. If the model does not describe its output in these terms, then robustness comes through increased uncertainty in the model output (wider confidence limits or thresholds) and reduced sensitivity in fault detection. Practically, the model training data have to be collected at one or two airflow rates. An improvement would be to retune the thresholds as new regions of operation are encountered, to maintain maximum sensitivity.

Table 7. Detection and Diagnosis of Faults During Spring Blind Test Period for AHU-A and AHU-B

Test Day	AHU A	AHU B	Physical Models: Detect from Innovations	Physical Models: Diagnose from Expert Rules	Physical Models: Diagnose from Recursive Parameter Estimation	Electrical Power Models: Detect	Electrical Power Models: Diagnose
1	Static pressure sensor offset (Stages 1-3)		Yes	No (unknown fault)	Yes	Yes	Yes
		Leaking cooling coil valve (Stage 1)	No	—	—	Yes	Yes
2	Unstable pressure control[1]		No fault[2]	—	—	No fault[2]	—
		Leaking cooling coil valve (Stage 2)	Yes	Yes	No[3]	Yes	Yes
3	Normal operation		No fault	—	—	No fault	—
		Leaking cooling coil valve (Stage 3)	Yes	Yes	Yes		—
4	Reduced cooling coil capacity (Stage 1)		Yes	No[4]	No[5]	Yes	Yes
		Normal operation	Yes	No[4]	—	No fault	—
5	Reduced cooling coil capacity (Stage 2)		No	—	No[6]	No	—
		Leaking recirculation damper (Stage 1)	No	—	—	No	—
6	Reduced cooling coil capacity (Stage 3)		Yes	Yes	Yes[7]	Yes	Yes
		Leaking recirculation damper (Stage 2)	No	—	No[6]	No	—
7	Unstable pressure control[1]		Yes	No[4,8]	No [8,9]	Yes	Yes
		Leaking recirculation damper (Stage 3)	Yes	Yes	No[6]	No	—
8	Unstable pressure control		Yes[10]	—	—	Yes	Yes
		Leaking recirculation damper (Stage 3)	Yes	No[11]	No[6]	No	—

[1]Implementation of unstable pressure control was considered by test implementers not to have taken effect.

[2]Although the day was labeled as normal by the physical model and electrical power FDD methods, periods of control instability were detected by both methods.

[3]Diagnosed as increased coil capacity (discharge-air temperature lower than predicted) or an unknown fault. An increase in coil capacity would not represent a degradation fault but could indicate that the cooling coil had been cleaned after the model calibration period.

[4]Diagnosed as a leaking cooling coil valve.

[5]Diagnosed as increased coil capacity.

[6]Diagnosed as an unknown fault in the mixing box.

[7]Alternative diagnosis as an unknown fault in the mixing box.

[8]The physical model FDD method detected brief periods of unstable control.

[9]Diagnosed as increased coil capacity.

[10]The unstable pressure controller was detected via the steady-state filter, which indicated that the measured pressure was in a dynamic state for a prolonged period.

[11]Diagnosed as an unknown fault.

Table 8. Detection and Diagnosis of Faults for AHU-1 Blind Test Period

Test Day	Fault	Physical Models: Detect from Innovations	Physical Models: Diagnose from Expert Rules and Recursive Parameter Estimation	Electrical Power Models: Detect	Electrical Power Models: Diagnose
July					
12	No fault	False alarm	Coil undercapacity	No fault	—
14	No fault	False alarm	False alarms from supply air and cooling coil models; design fault from mixing box model[1]	No fault	—
16	No fault	False alarm	False alarm from cooling coil model[2]	No fault	—
17	No fault	False alarm	False alarm from cooling coil model[2]	No fault	—
27	Aborted fault	No fault		No fault	
28	Static pressure sensor offset (Stages 1-2)	Yes	Yes	Yes	Yes
29	Static pressure sensor offset (Stages 2-3)	Yes	Yes	Yes	Yes
30	Stuck outdoor air damper	Yes[3]	Inconclusive diagnosis from mixing box model	No	—
31	Stuck outdoor air damper	No	—	No	—
August					
1	Normal operation	No fault	—	No fault	—
9	Cooling coil air-side fouling (Stage 1)	Yes[4]	False alarm from cooling coil model[2]	No	—
10	Cooling coil air-side fouling (Stage 2)	Yes[4]	False alarm from cooling coil model[2]	No	—
10-11	Cooling coil air-side fouling (Stage 3)	Yes[4]	False alarm from cooling coil model[2]	No	—
11-12	Leaking heating coil	Yes[5]	False alarm from supply air model; cooling coil capacity fault	No	—
13	Stuck-closed recirculation damper	Yes[5]	Yes—correct diagnosis from the supply air model; a false alarm from cooling coil model[2]	Yes	No
14, 15, 16	Normal operation	No fault	—	No fault	—
17	Loss of supply fan control, aborted after 2 h	Yes	No	Yes	No
18	Loss of supply fan control	No	—	No	—
19	Normal operation	No fault	—	No fault	—

[1]A fault was detected from innovations in each of the three models: supply air, mixing box, and cooling coil. The diagnosis from the supply air model was a false alarm due to a poor representation of the air distribution system. The diagnosis from the mixing box model was a design fault because the return fan control for AHU-1, a fixed percentage of the supply fan speed, influenced the proportion of outside air in the supply air. The diagnosis from the cooling coil model was a false alarm because the prediction of mixed air humidity needed in the model was influenced by the return fan control strategy.

[2]The diagnosis from the cooling coil model was a false alarm because the prediction of mixed air humidity needed in the model was influenced by the return fan control strategy.

[3]A fault was detected from innovations in the mixing box model.

[4]A fault was detected from innovations in the cooling coil model but the diagnosis was a false alarm.

[5]A fault was detected from innovations in the supply air and cooling coil models.

The lower levels of leakage in the cooling coil valve were difficult to detect. A significant factor that resulted in this insensitivity was the modeling of the valve. The nonlinear characteristics associated with the heat exchanger process combined with the poorly balanced chilled water circuit resulted in very high gain as the valve opens. This is difficult to model and results in a high degree of uncertainty in the low region of operation. It was found that leakage could only be detected when the valve was closed.

The leakage fault was not detected during the winter test period, which should have been the period when the fault was most visible. The steady-state detector deemed that an extremely high proportion of the data were transient and hence there were almost no data with which to monitor the system. This excessive dynamic activity in the coil control system was due to the cycling of the chilled water inlet temperature, due to the close coupling of the chiller to the coil and to the two-stage control of the reciprocating chiller. Although this could be considered to be a design fault, these effects were considered to represent "acceptable operation." The steady-state threshold was reset during the spring test period to allow more data through the steady-state detector and hence permit the fault monitoring function. The relaxed steady-state criterion, however, resulted in larger model prediction errors. The fault threshold were increased accordingly, again reducing the sensitivity of the method to detection.

The reduced coil capacity fault implemented on AHU-A and AHU-B was easily detected with the exception of spring operation, when only the highest level of the fault was detected. This was attributed to the relatively low load on the coil, which limited the effect of the fault on the coil performance.

The fan-duct reference model for the static pressure prediction was sufficiently accurate to allow the detection of the offset in static pressure, at least for the second and third magnitudes of the fault. The first two stages of the slipping fan belt were not detectable, although the observations from the data showed the effects on the performance to be very small. Oscillatory supply duct pressure control was detected by a prolonged period of dynamic activity (as indicated by a lack of steady-state data classified by the steady-state detector). This approach proved to be reliable: the fault was detected each time it was implemented.

The stuck-open outside air damper fault implemented on AHU-1 was not detected. The return fan overloaded the supply fan to such an extent that the relative proportion of the recirculation airflow increased from normal operation. The effect of the stuck-*open* outside damper was masked to the extent that the system appeared to have a fault similar to a stuck-*closed* outside air or exhaust air damper. The first-principles-based method indicated that a fault condition was present in the economizer on the day that the stuck-open damper fault was implemented, although no firm diagnosis could be made.

Neither the leaking heating coil valve nor the fouled cooling coil surface faults were detected in AHU-1. Both faults should have been detected by observation of the prediction error at the supply air temperature point, although it would have been unlikely that the faults could have been distinguished because they both result in a reduction in cooling coil capacity. Failure to detect both faults was due to the high degree of uncertainty in the model predictions. Factors contributing to this uncertainty that the point temperature sensors on the air-side are more susceptible to airflow-related temperature offsets such as stratification, and that the estimation of the mixed-air humidity ratio (inlet humidity to the cooling coil) was poor, partly because of the airflow imbalance problems discussed for AHU-A and B, and partly due to the estimation of the parameters of the economizer model from data taken from point temperature measurements. The cooling coil model was dehumidifying during the summer test period and model predictions were sensitive to uncertainty in the estimate of inlet humidity. These problems led to several false alarms during the AHU-1 test period. After completion of the test period, it was clear that these false alarms could have been avoided by a marginal increase in the level of the thresholds.

Conclusive diagnosis with both the expert-rules and the recursive parameter estimation was limited by the need for data to be available across the range of operation (low, high and mid ranges). For example, a leaking cooling coil valve could only be distinguished from a sensor offset if the fault was apparent only when the control valve was closed or nearly closed. However, during most of the tests implemented in this study, the systems remained in a narrow region of operation.

The problems associated with unmodeled disturbances, the lack of independence in the parameters and tests carried out over nominally one operating condition (season) led to difficulties in generating reliable performance from this method.

Gray-Box Correlations with Electrical Measurements

Results with submetered power data were very satisfactory for the three blind test periods for AHU-A and AHU-B. Almost all faults were detected. Careful maintenance and control of the HVAC systems and a limited pallet of faults to choose from made fault diagnosis possible, whereas it would be substantially more difficult or impossible in a less-controlled setting.

The stuck-closed recirculation damper was detected and diagnosed in the two test periods in which it was implemented. The leaky recirculation damper was the most difficult to detect. Analysis of chiller cycling frequency was limited to a narrow range of outdoor temperatures, to block the influence of outside temperature on chiller loading. Suitable conditions were present in the late-winter test and the fault was successfully detected and diagnosed. Temperatures were milder in the spring test and the fault was not found. A less restrictive temperature band, not evaluated, might have made it possible to find the fault in spring at the expense of possible false alarms.

The leaky cooling coil valve was detected and diagnosed in the two test periods in which it was implemented. The coil capacity fault was detected and diagnosed successfully in the summer test period and was also found on two of the three implementation days in the spring test period. It was not detected during the second of the three degradation stages in spring because the cooling loads were relatively low and the cooling coil valve did not open to an extent sufficient to reveal the fault.

The pressure sensor offset fault was detected and diagnosed successfully in all three test periods and the unstable fan controller was detected and diagnosed in the two periods in which it was implemented. All three degradation stages of the slipping fan belt were detected and diagnosed in the summer test period but only the most severe stage was found in the winter tests. At that time, the detection algorithm required that the fan-speed control signal be 100%, an unduly severe restriction that was met only on the last day, when there were large loads on the fan. The detection algorithm was changed for AHU-1, to rely on confidence intervals above and below the normal-operation correlation of fan power with speed, with no restriction on the speed signal as a prerequisite for detection of a fault.

As noted earlier, four of the six AHU-1 faults were entirely unknown to the investigators and had not been studied on AHU-A and AHU-B. The electrical power method successfully detected three of the six faults (stuck-closed recirculation damper, pressure sensor error, and loss of control of the supply fan), successfully diagnosed only one (pressure sensor error), and did not find the three remaining faults. Balancing this mixed performance, it is worth noting that one of the detected faults, the loss of control of the supply fan, was not among those for which the method had been commissioned. Further, the method did not generate any false alarms.

After the AHU-1 faults were revealed to the investigators, the electrical power FDD method was extended and applied with more care to data recorded during days when the undetected faults were implemented. The three faults still defied detection. Neither the stuck-open outside air damper nor the fouling on the cooling coil affected the supply fan power for a given airflow.

The impact of fouling on cooling coil capacity was not investigated because chiller cycling at high loads is strongly affected by unmeasured variables (internal and solar loads, for example). The leaking heating coil valve could not be detected via a change in power consumption of the source of hot water because the boiler was not monitored. An analogous method was successful in finding the leaking cooling coil valve, as already noted. While the leaking heating coil valve did introduce a heating load on the (downstream) cooling coil that affected the chiller cycling period, the change was not sufficiently conclusive to warrant flagging it as a fault.

DISCUSSION

All sensors required to implement the first-principles-based methods (Table 9) are typically installed in VAV systems for controlling the HVAC processes; the sole exception is the supply airflow sensor, which is not used when the return fan is controlled on the basis of supply fan speed. For effective fault isolation it is desirable to generate prediction errors at the outlet of each modeled subsystem. The attendant sensor is required to make this possible. Improvements to the performance of the cooling coil FDD scheme could have been realized if the coil outlet air temperature measurement was used, rather than the supply air temperature (the fan temperature rise model would not have been necessary).

Results from the test periods demonstrated that the accuracy of the cooling coil model predictions under dehumidifying conditions might have been improved if a better estimate of the inlet humidity ratio were available. A possible solution is the installation of an additional sensor if the cooling coil is designed for latent duty, although the air local to the sensor location needs to be well mixed. An alternative would be to measure the ambient airflow rate as well as the supply airflow rate. The proportions of ambient and return air in the supply air could then be calculated directly.

Ideally, the outside-air temperature and humidity sensor would be located in the inlet duct to the system rather than being external to the building to reduce the effects of improper representation of the air properties. This would inevitably increase the cost of the implementation of the FDD scheme.

The electrical power FDD method required substantially fewer measurements, which reduced the sensor maintenance requirements. Sensors for this method are listed in Table 9. The electrical power data that are at the heart of the method are not available in typical HVAC plants. The method as implemented in the test building required a power meter for the supply fan for each AHU. Another power meter was required for the single onsite chiller. The apparent economy in having just one chiller and one chiller power meter was more than outweighed by difficulties in ascribing changes in chiller cycling to faults in a particular air handler. To provide an informative test of the electrical power FDD method, it was necessary to couple the chiller to a single air handler and use district chilled water for the others, even when cooling loads were low and the single chiller would have had adequate capacity. In a high-rise office building where a single chiller serves separate air handlers on each floor, the electrical power method would not be able to detect faults on the basis of chiller power.

The sensors available for use on this project were generally instrument-grade devices having a higher accuracy, more stability and less drift than standard HVAC-grade devices. Exceptions include the return air humidity sensor and the supply duct static pressure transducers, which were standard HVAC-grade devices. Equipment costs for instrument-grade sensors are typically 3 to 5 times the cost of standard, mass-produced HVAC-grade sensors. Costs noted below exclude installation, setup, and any onsite calibration, which is estimated to average $75 per sensor.

Sensors that require periodic calibration, such as those measuring temperature, pressure, and flow, were within their appropriate calibration dates to comply with certification requirements.

Table 9. Sensors and Control Signals Required for Implementation of Each FDD Method

Sensor Type or Control Signal	Condition	First-Principles-Model FDD	Electrical Power FDD
Temperature	Return air	X	
	Outside (ambient) air	X	X
	Mixed air	X	
	Supply air	X	
	Chilled water flow to coil	X	
Humidity	Return air	X	
	Outside (ambient) air	X	
Flow	Supply air	X	X
Pressure	Supply duct static pressure	X	X (training only)
Electrical power	Chiller		X
	Supply fan		X
	Secondary chilled water pump		X
Control Signal	Return fan	X	
	Economizer	X	
	Cooling coil control valve	X	X
	Supply fan	X	X

Note: All sensors and control signals are required for fault detection; there would no reduction in sensor count if monitoring were limited to detection and excluded diagnosis.

Calibration represents the most significant maintenance requirement for the sensors. Typically, the listed sensors are calibrated annually, and more frequently than sensors in a standard HVAC system:

- The return, supply, and outside airflow rates and temperatures were measured by electronic airflow measuring stations. Each measuring station costs $1400 and has a stated accuracy of ±2% for airflow greater than 2.54 m/s (500 fpm) and ±0.2°C (0.36°F) for temperature, with zero long-term drift.
- Mixed-air temperatures were sensed using instrument-grade 1000-ohm platinum RTDs arranged in a multipoint array. The array has a listed device accuracy of ±0.14°C (0.25°F) and an average cost of $2500. Return air humidity is measured with a standard HVAC-grade humidity sensor with an accuracy of ±3% and an equipment cost of $50.
- Supply duct static pressure transducers were standard HVAC-grade devices with a stated accuracy of ±1% of full scale [±0.75 Pa (0.03 in. of water)]. Stability was listed as ±1.0% of full-scale deviation from original calibration for one year under normal operating conditions. These transducers cost $225 each, compared with $485 for sensors with an accuracy of ±0.25% of full scale, which were installed after this project was completed.
- Water temperatures were sensed with single point, direct immersion, instrumentation-grade, 1000-ohm 2-wire and 100-ohm 4-wire, platinum RTD sensors. The sensors have a stated accuracy of ±0.14°C (0.25°F) and are very stable, with little long-term drift. These temperature sensors and related equipment have an average cost of $125 each.
- Electrical power to the supply fans, pumps and chiller was measured with precision AC watt transducers with a stated accuracy of ±0.2%. These devices have NIST-traceable calibration. The watt transducers cost $400 each, including $60 for calibration; the transducer for the chiller required additional current transducers and cost $550. HVAC-grade watt transducers,

not used at this site, are available for about $250.

- Outside air conditions were measured by an instrument-grade temperature and humidity measuring weather station. A 100-ohm platinum RTD with a stated accuracy of ±0.2°C (0.36°F) was used for temperature and a polymer sensor with a stated accuracy of ±1% (at 0 to 90%rh) was used for humidity measurement. The measuring station has a total cost of approximately $1000.

The overall accuracy of the instrumentation system depends on factors in addition to the stated accuracy of the sensing device: transducers, stability of power supplies, wiring type, lead length, A/D converters, processor resolution, scaling values and, in particular, measurement representation of average fluid properties and quantities. The impact of these uncertainties on the performance of FDD methods was not evaluated as part of this work. A full assessment of the uncertainties associated with the first-principles-based FDD methods in HVAC systems is presented in Buswell (2001).

The parameters of the first-principles models need to be calibrated for each test system. Some parameters are identified directly from design information and/or inspection. The remaining parameters are simultaneously identified using test data from the target system. These data were obtained by increasing the control signal to each sub-subsystem in a series of steps from 0% to 100% and back to 0%. Each step is held until steady state is considered to exist. For the tests conducted in this research, the total time taken to commission three subsystems in one AHU (economizer, cooling coil, and fan-duct system) was 23 h. This was controlled largely by the time constants associated with the system and the need for the observation of a number of consecutive points at each step to decide whether steady-state conditions exist. To decrease this overhead, simultaneous commissioning of all three subsystems was investigated on AHU-1 (Norford et al. 2000). The simultaneous test took 14 h to complete, saving 9 h. Alternative schemes to generate training data may provide a quicker method of capturing the system characteristics.

The model parameters are described fully in Norford et al. (2000) and are listed in Table 10. The commissioning tests were designed to provide data to capture the principal system characteristics. It was not always possible to identify all parameters simultaneously such that each parameter represented its prescribed system characteristic. This was caused by a lack of parameter independence, which increased as the number of parameters increased (i.e., one parameter estimate may in part be offset by the value in another and hence the subsequent parameter estimates become gray).

This phenomenon, particularly apparent in the economizer model, led to an approach in which subsets of the parameters were estimated from the commissioning data that most related to their effect on the model. Leakage parameters were identified from data for which the control elements were closed (0% or 100% control signal). With the leakage parameters fixed, the model "gain" parameters were then identified from the data for the opening movement of the control signal (0% control signal to 100% control signal). Finally, with the leakage and gain parameters fixed, the control actuator hysteresis parameters were identified from all of the commissioning data (0% control signal to 100% control signal and then reverse from 100% to 0%).

The steady-state detector also required two parameters for each monitored subsystem. One describes the dominant subsystem time constant in relation to system dynamics and the other controls the amount of data that is considered to be transient. As the stringency of the parameter is increased, the amount of data passed on to the FDD methods is reduced. Trials on a number of HVAC systems, including those systems tested here, have revealed that the values of these remain quite consistent for similar subsystems.

Table 10. Parameters of First-Principles Models

Subsystem	Design Parameters	Calibrated Parameters
Fan-duct	Rotational speed of the fan (estimated from the control signal) (–)	*Static pressure sensor offset (Pa)*
	Static pressure at zero mass flow rate (Pa)	Total fan/duct resistance to airflow (sPa/kg)
	Fan minimum rotational speed (–)	
	Fan maximum rotational speed (–)	
	Control signal relating to minimum rotational speed (–)	
	Speed at which fan belt slippage occurs (–)	
Air temperature rise due to fan	Control signal relating to the minimum fan speed (–)	Minimum temperature rise (K)
		Maximum temperature rise (K)
Economizer (dampers)		Mixed air temperature offset (K)
		Parameter that describes the degree of curvature in the process relationship (–)
		Parameter that defines the asymmetry of the process (–)
		Leakage through the return damper (–)
		Leakage through the outside air damper (–)
(actuator)		High activation point (–)
		Low activation point (–)
		Hysteresis (–)
Cooling Coil (coil)	Coil face area (m^2)	*Heat-transfer scaling factor (–)*
	Number of rows (–)	Supply air temperature sensor offset (K)
	Number of circuits for parallel flow (–)	*Fractional flow leakage (–)*
(three port control valve)	Maximum chilled water mass flow rate (kg/s)	Curvature coefficient (–)
		Authority (–)
(actuator)		High activation point (–)
		Low activation point (–)
		Hysteresis (–)

Note: The parameters in italics are those used for fault diagnosis.

In concept, the methods do not require the selection of fault thresholds. A fault is simply detected when a prediction error is considered to be significant against some statistical measure. Similarly, a statistically significant change in the value of a recursively re-estimated fault parameter could also be used to indicate the presence of a fault. In practice, the scheme did not formally account for the uncertainty with respect to the sensitivity of the model output to unmodeled phenomena. The addition of the thresholds, above which a prediction error generated an alarm, was a subjective and somewhat ad hoc attempt to account for this uncertainty. Better model representation of the process and a formal methodology to account for the uncertainties

from all sources has been shown to produce a more sensitive and a more robust first-principles-based fault detection scheme than that used in this research (Buswell 2001).

The expert-rule method that employs the bins requires a number of parameters to locate the bins within the operating space for each subsystem (i.e., what value of control signal is attributed to low, mid and high operation). The expert rules are generic and principally describe what evidence is expected to exist with respect to the fault and the bin categories.

The sensitivity of the electrical power FDD method to faults and robustness with respect to false alarms are functions of the extent to which the semiempirical power correlations capture the process characteristics and the calculation of a large number of thresholds and other parameters, listed in Table 11. These thresholds and parameters can be usefully grouped as follows: (1) normal equipment power levels, from measurement or manufacturer's data; (2) statistical confidence intervals; (3) fault detection thresholds based on commissioning the FDD method with known faults; or (4) data analysis regions, based on commissioning the FDD method with known faults and designed to improve the robustness of the method in both detection and diagnosis.

The list of parameters does not include the parameters in the third-order polynomial curve fits that express the correlation of electrical power with flow or rotational speed. Values for these parameters were determined from training data for individual fans and pumps.

Required values for the first two groups are easily obtained or assigned. One goal of ongoing research is to reduce the number of parameters and thresholds in the last two groups, to simplify and eventually automate the process of commissioning this FDD method. To that end, the AHU-1 test period in this research provided an opportunity to replace the effective but hand-tuned analyses of fan power as a function of speed and pump power as a function of valve position with more straightforward polynomial power correlations, for which the statistics are rigorous and the only need is to supply a confidence interval.

Calibration of the power correlations for normal operation was entirely a passive procedure and demanded only a reasonable range of operating conditions. For the test building, about 10 h of data from power meters and supporting sensors were required. At the test building, these data were collected in a single summer season. These data were sufficient for fault detection but not diagnosis, where knowledge of fault signatures was required to distinguish faults that reveal themselves as a deviation in a power correlation for a single component (i.e., the supply fan). Fault diagnosis rules developed for equipment in the test building relied on observations of faulty performance. Further tests in different buildings are necessary to determine the extent to which these rules are general or can be easily adjusted. It is clearly not practical to commission an FDD method with onsite faults.

The first-principles-based methods are formulated to represent what is normally considered as "ideal" system operation. In this respect, they are sensitive to any nonideal system behavior, which could represent a design fault. Two forms of nonideal behavior affected the sensitivity and robustness of the first-principles methods during these tests. Both forms concerned changes in the relative proportion of outside and recirculation airflow rates through the mixing box, as the supply fan speed varied. One form of this phenomenon is reported in Seem et al. (1998). The nonideal behavior could cause innovations in both the mixed-air temperature and supply air temperature, suggesting faults in the economizer and cooling coil respectively.

In addition to the effect of the nonideal system behavior, the first-principles-based methods were sensitive to the changeover in system configuration necessary at the start of each test period (the test systems were used for other project work in between the seasonal tests conducted in this study). The changeover in configuration concerned control strategies and reinstallation of some sensors. Further, some physical disruption to the systems, such as repairs to

Table 11. Thresholds and Other Parameters Required by Electrical Power FDD Method

Description of Thresholds and Parameters	Value
Fan-power correlations with airflow and speed control signal	
Maximum deviation of static pressure from set point for training data	25 Pa (0.1 in. of water)
Confidence level to establish boundary between normal and faulty data	90%
Airflow boundary to distinguish stuck-closed recirculation damper from static pressure offset/drift[1]	500 cfm
Fan power at 100% speed below which a slipping-fan-belt fault was flagged, subject to a minimum time duration[2]	1 kW
Time duration for low fan power at 100% speed, above which a slipping-fan-belt fault was flagged	3 one-min. power samples
Pump-power correlation with cooling coil valve position control signal	
Valve position control signal above which pump-power data were analyzed for a cooling coil capacity fault[3]	40%
Measured normal-operation power level of the secondary chilled water pump	400 W
Minimum decrease of pump power below normal-operation value, in excess of which a coil capacity fault was flagged[4]	10 W
Confidence level to establish boundary between normal and faulty data (used for AHU-1)	90%
Chiller-cycling analysis	
Power level above which the chiller is considered to be operating in the low-power stage[5]	4 kW
Cycling interval when the cooling coil valve control signal is at 0%, below which a leaky-valve fault is flagged[4]	30 min.
Normalized outdoor air temperature, below which chiller cycling is analyzed to detect a leaky recirculation damper[6]	0.2
Power-oscillation analysis	
Size of sliding window for averaging one-minute power data from submeters	5 samples
Standard deviation of power signal above which a fault is flagged, as a percentage of average power	15%

[1]This parameter was used solely for fault diagnosis.

[2]Fan-power analysis at 100% speed was used in AHU-A and B to detect the slipping fan belt. For AHU-1 this approach was replaced by the more rigorous and sensitive polynomial correlation of fan power with speed control signal.

[3]Pump-power analysis relative to a measured and near-constant normal-operation value was used in AHU-A and B to detect the coil capacity fault. For AHU-1 this approach was replaced with a polynomial correlation of pump power with valve position control signal.

[4]The parameters for the decrease in pump power and the change in chiller cycling interval were used for a single-stage detection and diagnosis in the test building, where the number of faults was limited.

[5]The chiller's high-power stage was not of concern, because the chiller-cycling analysis was limited to low-load conditions when the chiller was either off or in the low-power state.

[6]The normalized outdoor air temperature, defined in Shaw et al. (2002), is the difference between the outdoor air temperature and the supply air temperature set point, normalized by the difference between the supply and room air temperature set points.

damper linkages, was necessary between test periods. This could have caused a change in the system characteristics and reduced the validity of the model calibration.

The nonideal system characteristics and the necessary disruption to the state of the system that occurred between test periods resulted in the setting of fault thresholds for each test period. This was done in a subjective manner to eliminate false alarms during periods of

known normal operation (Norford et al. 2000). This approach was not completely effective for the AHU-1 tests, during which the fault thresholds were adjusted for two consecutive days of operation. Following the setting of the thresholds, the nonideal system behavior became more predominant, which led to several false alarms for the test period. Although it was recognized that the alarms were false and that they could be eliminated with very minor changes to the thresholds (~0.25°C), the thresholds were not readjusted during the fault testing, because this would not have been consistent with the FDD methodology as proposed for this project. In practice, the thresholds could easily be made more robust by setting them over a longer period of operation.

In some instances, the nonideal system behavior resulted in the setting of relatively wide fault thresholds, which necessarily reduced the sensitivity of the methods, particularly for the more subtle leakage faults. Considering that the systems were subject to change between test periods and that they exhibited nonideal behavior, the results of this study make it clear, however, that the first-principles methods remained robust in fault detection. The models were not recalibrated to account for disruption to the systems that occurred between test periods.

The diagnosis of faults by the first-principles methods was less robust than the fault detection. A number of faults detected during the blind test periods were misdiagnosed. Robust diagnosis by expert rules requires the system to have operated over its complete range during the occurrence of a fault. Considering that some faults can force the system to move to, and remain at, one operating point, this requirement is impractical. However, this restriction could be eliminated by developing a methodology that includes the injection of test signals to exercise the system across its range of operation once a fault has been detected.

Fault diagnosis by the recursive re-estimation of the first-principles-based model parameters was sensitive to the unmodeled disturbances, the limited excitation in terms of operating condition, and the lack of independence of the parameters. It is unlikely that it would be possible to include parameters to specifically represent all fault conditions. Evidence from this work suggests that typical HVAC system data could probably support two recursively re-estimated parameters; one describing the under/over capacity at the "high duty end" of operation and one describing under/over capacity at the "low duty end" of operation.

The electrical power FDD method was effective in detecting faults for AHU-A and AHU-B. The detection methodology was straightforward for some faults, including those that affected fan power for a given airflow (the stuck-closed recirculation damper and the pressure sensor offset), those that impacted the cycling of the reciprocating chiller, and the oscillating controller. The detection methodologies for other faults were developed in response to the configuration and operation of the equipment in the test building. Notably, the cooling coil capacity fault was detected by identifying changes in the electrical power drawn by the secondary chilled water pump. This method was effective only because the fault was introduced in a way that significantly obstructed the water flow. Water-side tube fouling would likely not have been detected. Detection of the slipping fan belt was effective but done with a threshold developed from observations of the impact of the fault, rather than by relying strictly on a fan power correlation. Detection of this fault for AHU-A and AHU-B was also limited to 100% speed control signal, an apparently unnecessary restriction on the use of the power-speed correlation that was eliminated for AHU-1. This same restriction was employed in the first-principles model to distinguish this fault from the pressure sensor offset.

Several of the faults introduced into AHU-1 did not produce detectable changes in electrical power, including the stuck-open outside air damper, the obstruction of the cooling coil, and the leaky heating coil valve. Detection failure in these cases is acceptable, because the FDD method developed for other faults was rationally applied and simply did not reveal significant deviations from normal operation. It is possible that further tuning of the method could have reliably

revealed the leaky heating coil valve, in the test building and others where the heating coil is upstream of the cooling coil. This will be pursued in the future.

As with the first-principles method, fault diagnosis via the electrical power FDD method was less reliable than fault detection. Diagnosis techniques also required considerably more effort to develop and commission. For example, it was relatively easy to measure the cycling periods of the reciprocating chiller, but considerable care was required to establish operating regions where the leaky recirculation damper could be diagnosed. Inadequate care in this step not only confounds the diagnosis process but leads to false alarms, as would occur if a change in chiller cycling were flagged but that change were due solely to variations in the thermal load on the chiller. Diagnosis of the stuck-closed recirculation damper and the pressure sensor offset required careful observation of the HVAC plant and a knowledge of the normal control of the mixing box dampers at different times of day and different seasons.

The electrical power method proved to be robust in terms of false alarm generation. None were generated in the AHU-1 test period. In part, this is due to the choice of confidence intervals for the power correlation; intervals of lower confidence would be tighter and would tend to reveal more faults and generate false alarms.

CONCLUSIONS

The relatively rare opportunity to thoroughly test FDD methods in a building operated as a research facility is an invaluable bridge between simulation and lab testing and field deployment. Such "real-world" issues as sensor placement and calibration, fault magnitudes, and imperfectly understood equipment performance under normal and faulty operation make it unlikely that a first-generation FDD method will successfully leap from lab to commercial use. Controlled field tests such as were required for this project are not so much a proof of performance of fully mature methods but a means of revealing flaws in the methods and subsequently refining them to the point where their performance is substantially improved.

The first-principles-model-based methods can be implemented without installation of special sensors. However, further work is required to shorten the time required to gather the training data for model calibration. The electrical power correlation method requires the use of additional electrical power sensors, but models are calibrated primarily from data collected during normal system operation. The advantage of a single testing procedure and models that can extrapolate is that the FDD scheme is operational immediately after installation (and calibration) of the FDD software. Normal operation data could be used to calibrate the first-principles-based models in a similar manner to the electrical power correlation approach. The disadvantage, however, is that data from across the subsystem operating season are required before the models are fully calibrated.

As shown in Tables 5 through 8, both FDD methods performed reasonably well in detecting faults. The electrical power FDD method was less developed and may not apply to other sites. Power correlations can be used as an effective method for detecting faults, but it is not clear how much work will be required to adapt the basic approach to different plants. The first-principles-based scheme is more mature than the electrical power correlation scheme as an FDD approach. The methods have been applied to other systems installed in real buildings. The reduction of false alarms and increase in sensitivity in the detection of faults has been addressed in subsequent work by Buswell (2001). In the future, developers of the electrical power method will consider extending that method to use additional sensors, such as temperature sensors typically found in control systems. Table 12 summarizes the evaluation of the two methods.

Both FDD approaches require some additional information to diagnose faults. This effort was simplified at the test site because of the limited number of introduced faults. Implementation of

Table 12. Comparison of First-Principles Physical Models FDD Method and Gray-Box Electrical Power Method

Feature	First-Principles Physical Models	Gray-Box Electrical Power Models
Operates only on steady-state data	Yes	Yes
Calibration time	15 to 23 h	10 h
Active or passive training data	Active; a single training period is sufficient	Passive; need seasonal data
Training methodology	Well defined	Less well defined
Thresholds	Statistical confidence intervals, arbitrarily selected thresholds	Statistical confidence intervals, arbitrarily selected thresholds
Critical sensors	Supply, return, mixed-air and outdoor air temperature sensors, return air and outdoor air humidity, supply airflow, control signals	Electrical power meters, supply airflow, control signals for valves and fan variable-speed drives
Fault detection	Excellent results for faults which the methods were tuned to find	Excellent results for faults which the methods were tuned to find
Fault diagnosis	Moderate number of misdiagnoses	Good results when list of potential faults is small
False alarms	Moderate	None
Extension of method to larger list of faults	Possible	Some faults cannot be detected with electrical power measurements
Other issues	Excessive number of fault parameters in models	Effects of faults on electrical power must be carefully defined for each component (fan, pump and chiller) in each system (constant volume, VAV, etc.)

either method on a system with no prior knowledge of the causal faults will almost certainly result in ambiguous diagnosis.

The first-principles-based methods are sensitive to the occurrence of nonideal system behavior. The detection of nonideal system behavior by the method is an advantage where nonideal behavior is considered to be a design fault. However, where the nonideal behavior must be accepted as part of the system characteristic, the uncertainty in the prediction error increases, reducing the sensitivity of the method to fault detection. Better modeling of the system behavior results in a reduction in the uncertainty in the model predictions and improved fault detection rate. The gray-box electrical power models are less sensitive to nonideal system behavior because the correlations model the system behavior under closed-loop control. The gray-box modeling methods were, therefore, more robust than the first-principles-based methods in that they generated no false alarms.

Robust fault diagnosis using first-principles-based models and expert rules is limited by the need for the system to have moved across its range of operation during the occurrence of a fault condition. A way of generating this information "out of season" would be to inject test signals

that excite the system across the deficient region. Fault diagnosis by the recursive re-estimation of model parameters did not prove to be reliable in the test environment in this research.

The apparent complexity of both methods reflects their degree of development. Much of the identification and data gathering processes described could be fully automated. As the methods mature, the selection of some of the required parameters will become better understood and easier. There are still issues surrounding the sensitivity and robustness of FDD methods (Dexter and Pakanen 2001). A balance between sensitivity in detecting faults and robustness in minimizing false alarms is needed. More testing with data sets from real buildings is required, and in particular a concerted effort is required to generate reliable, streamlined, and automated commissioning processes for FDD methods. Both FDD methods investigated here need to be simpler in terms of application and the analysis of the data and need to be more "transparent" to the end user.

ACKNOWLEDGMENT

The authors warmly acknowledge the financial support of ASHRAE, the California Energy Commission (via subcontracts from Lawrence Berkeley National Laboratory and Architectural Energy Corporation), and Honeywell, Inc.; the vision of Phil Haves, who played a leading role in conceiving this project; and the technical advice of the ASHRAE Project Monitoring Committee, chaired by John House.

REFERENCES

Buswell, R.A. 2001. Uncertainty in the First Principle Model Based Condition Monitoring of HVAC Systems. Ph.D. thesis, Loughborough University, U.K.

Dexter, A.L. and M. Benouarets. 1996. A Generic Approach to Identifying Faults in HVAC Plant. *ASHRAE Transactions* 102(1):550-556.

Dexter, A.L. and J. Pakanen, eds. 2001. *Demonstrating Automated Fault Detection and Diagnosis Methods in Real Buildings*. IEA Annex 34 Final Report. VTT Building Technology, Espoo, Finland.

Haves, P., T.I. Salsbury, and J.A. Wright. 1996. Condition Monitoring in HVAC Subsystems Using First-Principles Models. *ASHRAE Transactions* 102(1):519-527.

Hyvarinen, J. and S. Karki, eds. 1996. *Building Optimization and Fault Diagnosis Source Book*. Technical Research Center (VTT), Espoo, Finland

Lee, W.Y., C. Park, and G.E. Kelly. 1996a. Fault Detection of an Air-Handling Unit Using Residual and Parameter Identification Methods. *ASHRAE Transactions* 102(1):528-539.

Lee, W.Y., J.M. House, C. Park, and G.E. Kelly. 1996b. Fault Diagnosis of an Air-Handling Unit Using Artificial Neural Networks. *ASHRAE Transactions* 102(1):540-549.

Norford, L.K., J.A. Wright, R. Buswell, and D. Luo. 2000. Demonstration of Fault Detection and Diagnosis Methods in a Real Building (ASHRAE 1020-RP) ASHRAE 1020-RP Final Report.

Peitsman, H.C. and V.E. Bakker. 1996. Application of Black-Box Models to HVAC Systems for Fault Detection. *ASHRAE Transactions* 102(1):628-640.

Salsbury, T.I. 1996. Fault Detection and Diagnosis in HVAC Systems Using Analytical Models. Ph.D. thesis, Loughborough University, U.K.

Seem, J.E., J.M. House, and C.J. Klaassen. 1998. Volume Matching Control: Leave the Outdoor Air Damper Wide Open. *ASHRAE Journal* 40(2):58-60.

Shaw, S.R., D. Luo, L.K. Norford, and S.B. Leeb. 2002. Detection of HVAC Faults via Electrical Load Monitoring. *International Journal of HVAC&R Research* 8(1).

Stephan, W. 1994. Comparison of Different Models for Cooling Coils Under Wet Conditions. *Proceedings of the Fourth International Conference on System Simulation in Buildings, Liege,* p. 291.

Yoshida, H., T. Iwami, H. Yuzawa, and M. Susuki. 1996. Typical Faults of Air-Conditioning Systems and Fault Detection by ARX Model and Extended Kalman Filter. *ASHRAE Transactions* 102(1):557-564.

California Energy Commission
Public Interest Energy Research Program

HPCBS # E5P22T3d

HPCBS

High Performance Commercial Building Systems

Advanced Nonintrusive Monitoring of Electric Loads

Element 5 - Integrated Commissioning and Diagnostics
Project 2.2 - Monitoring and Commissioning of Existing Buildings

Massachusetts Institute of Technology

IEEE Power and Energy
Magazine, March/April, 2003. 56-63.

pier
PUBLIC INTEREST ENERGY RESEARCH
"Research Powers the Future"

BERKELEY LAB

United States Department of Energy
Science, Security and Energy: Powering the 21st Century

Acknowledgement

This work was supported by the California Energy Commission, Public Interest Energy Research Program, under Contract No. 400-99-012 and by the Assistant Secretary for Energy Efficiency and Renewable Energy, Building Technologies Program of the U.S. Department of Energy under Contract No. DE-AC03-76SF00098.

Advanced Nonintrusive Monitoring of Electric Loads

Christopher Laughman Douglas Lee Robert Cox
Steven Shaw Steven Leeb Les Norford Peter Armstrong

I. Introduction

THE boom in communications research and development has made highly sophisticated means for gathering, moving, and exchanging information part of everyday life. Examples include the Internet, Bluetooth and 802.11b networks, cellular, and satellite communications. However, the process of gathering and analyzing the data to send over this expanding web of information networks remains expensive for many applications. A building facilities manager may be able to configure a cellular phone to monitor the energy usage and operating schedule of an HVAC plant, but the usefulness of the information transmitted will generally be directly proportional to the complexity and size of the installed sensor array.

This difference between information gathering and information networking serves to re-emphasize an established fact: though remote access to information and control inputs may be obtained easily and inexpensively via networking, access does not provide useful information without installation of a potentially expensive and intrusive sensor array. Mass production may ultimately reduce sensor cost, especially for solid-state or technologically advanced micro-electromechanical sensors. Installation expenses are likely to remain high, especially for temporary monitoring applications where data is gathered for a brief window of time with a removable sensor network. In addition, the reliability of a monitoring system with many sensors may be reduced in comparison to a system with relatively fewer sensors.

The Nonintrusive Load Monitor (NILM) can determine the operating schedule of electrical loads in a target system from measurements made at a centralized location, such as the electric utility service entry. In contrast to other systems, the NILM reduces sensor cost by using relatively few sensors. The NILM disaggregates and reports the operation of individual electrical loads like lights and motors using only measurements of the voltage and aggregate current at the utility service point of a building. It can also identify the operation of electromechanical devices in other kinds of power distribution networks. For example, the NILM can determine the load schedule in an aircraft from measurements made only at the generator, or in an automobile from measurements made at the alternator/battery block. The NILM can distinguish loads even when many are operating at one time.

The NILM is an ideal platform for extracting useful information about any system that uses electromechanical devices. The NILM has a low installation cost and high reliability because it uses a bare minimum of sensors. It is possible to use modern state and parameter estimation algorithms to verify remotely the "health" of electromechanical loads by using a NILM to associate measured waveforms with the operation of individual loads. The NILM can also monitor the operation of the electrical distribution system itself, identifying situations where two or more otherwise healthy loads interfere with each other's operation through voltage waveform distortion or power quality problems.

Strategies for nonintrusive monitoring have developed over the last twenty years. Advances in computing technology make a new wealth of computational tools useful in practical, field-based NILM systems. This paper reviews techniques for high performance nonintrusive load and diagnostic monitoring and illustrates key points with results from field tests.

II. Background and Early Approaches

One of the earliest approaches to nonintrusive monitoring, developed in the 1980's at MIT by Professor Fred Schweppe and Dr. George Hart, had its origins in load monitoring for residential buildings. Under Dr. Hart's ingenious scheme, the operating schedules of individual loads or groups of loads are determined by identifying times at which electrical power measurements change from one nearly constant (steady-state) value to another. These steady-state changes, known as events, nominally correspond to the load either turning on or turning off, and are characterized

by their magnitude and sign in real and reactive power. Recorded events with equal magnitudes and opposite signs are paired to establish the operating cycles and energy consumption of individual appliances. This process of detecting steady-state changes provides the basis for a commercial version of this early work, as described in a 1999 article in the IEEE Computer Applications in Power.

The 1999 CAP article [2] describes a five-step process for load disaggregation through the detection of changes in aggregate power consumption. First, an edge detector is used to identify changes in steady-state levels. Second, a cluster analysis algorithm is used to locate these changes in a two-dimensional "signature space" of real and reactive power (a ΔP-ΔQ plane). The signature space reduces the potentially complicated load transient data to a two-dimensional space of changes in power consumption with a pleasing and useful graphical interpretation. Third, positive and negative clusters of similar magnitude are paired or matched (especially for "two-state" loads that turn on, consume a relatively fixed power, and turn off). In a fourth step known as anomaly resolution, unmatched clusters and events are paired or associated with existing or new clusters according to a best likelihood algorithm. In the fifth and final step, pairs of clusters are associated with known load power consumption levels to determine the operating schedule of individual loads. This step uses information gathered during a training or survey phase in the building.

This five-step approach to nonintrusive load monitoring has the advantage of making intuitive sense, and has been demonstrably successful in certain classes of buildings, such as residences. Recent field tests in a variety of other buildings have uncovered a number of limitations of the technique of examining steady-state changes in a signature space. Some of these limitations are well established, while others are relatively new.

The two-dimensional signature-space technique relies on at least three key assumptions that limit its effectiveness. The first assumption is that different loads of interest exhibit unique signatures in the ΔP-ΔQ plane. They may not, especially in commercial and industrial facilities where the number and variety of loads is generally greater than in a residence. The two-dimensional signature-space becomes crowded with indistinguishable loads as the number and kind of loads increases. In a home there may be only one 5 kW load, such as a hot water heater. In comparison, a light commercial facility with an HVAC plant, office equipment, and lighting may have several different loads that overlap ambiguously in the ΔP-ΔQ plane.

The second limiting assumption is that load composition is determined by steady-state power consumption. This requires waiting until transient behavior settles out so that steady state values can be measured. We have found it difficult to find any suitable time scale in industrial or commercial environments that yields reliable steady-state measurements. A conservative estimate of steady-state power might require a long interval of near-constant demand. Requiring a long steady-state waiting time prevents the monitor from tracking rapid sequences of loads. Some loads will be missed, and these may not be caught in the anomaly resolution phase. Short waiting times, on the other hand, may trigger measurements in the middle of load transients, resulting in spurious events in the cluster analysis phase. Field studies have shown that large HVAC loads such as fans and chillers might take from 30 seconds to several minutes to gradually spin-up to their final operating speed. Other loads, such as variable speed drives (VSDs), may never settle to a steady-state. Furthermore, if the variations in power consumption are large enough, VSDs and similar loads could prevent the monitor from finding a steady-state consumption level or recording any events. Some industrial loads that ordinarily settle to steady state conditions may fail to do so if they include poorly tuned controllers or other faults.

A third limitation is that most steady-state NILM systems process data in "batch" format using a day or more of stored data. This is not strictly implied by the five-step disaggregation procedure outlined above, but is based on the assumption that near-real-time identification of load operation is not necessary. This limits the monitor to load survey and power scorekeeping applications, excluding the vast potential for applications in real-time diagnostics.

III. Advanced Techniques for Monitoring

Steady-state monitoring techniques are successful in homes and small businesses because of the low event generation rate and number of loads at these sites. Medium to large size commercial and industrial facilities require a more sophisticated approach, due in part to high rates of event generation, load balancing, and power factor correction. Savvy commercial and industrial facilities managers want near-real-time monitoring information in addition to batch results over an interval of several weeks. Real-time identification is also essential if the NILM is to serve as a "black-box" diagnostic monitor in transportation systems like aircraft or naval vessels. Field tests have demonstrated the value of a number of enhancements that enable an advanced NILM to tackle complex monitoring environments.

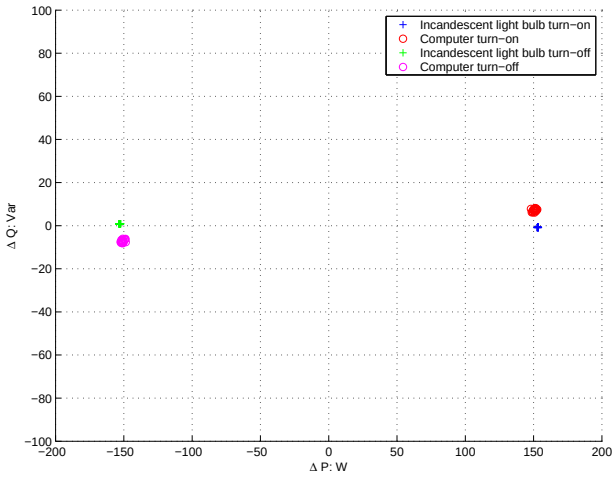


Fig. 1. Plot of the steady state power consumption of a computer and a bank of incandescent lights. Real power is plotted against reactive power.

A. *Higher Harmonics*

Higher harmonics in the aggregate current signal can be used to distinguish loads with overlapping clusters in the ΔP-ΔQ signature space. Many loads draw distorted, non-sinusoidal current due to their inherent physical characteristics or the presence of power electronics. Examples include office equipment (i.e., computers and copiers), actively controlled industrial equipment such as variable speed fans, and electroplating baths. Our advanced field monitoring system uses a phase-locked short-time Fourier transform of current waveforms collected at sample rates of 8000 Hz or higher to compute "spectral envelopes" that summarize time-varying harmonic content. For a single phase load, real and reactive power correspond to the envelopes of in-phase and quadrature current drawn by the load relative to the voltage. The short-time Fourier transform computes estimates of the real, reactive and higher frequency components of the current.

Figure 1 shows the that higher harmonics are useful for distinguishing loads that are similar in the P-Q signature space. Figure 1 shows a collection of turn-on and turn-off events recorded at a site with personal computers and a bank of incandescent lamps. The loads are practically indistinguishable in the P-Q signature space because they consume essentially the same real and reactive power. However, typical computer power supplies draw "signature" third harmonic currents. The loads are easily separable by examining higher harmonics.

Figures 2 and 3 illustrates the relative ease of distinguishing individual loads in a three-dimensional space with axes denoting changes in real power, reactive power, and third harmonic. Our advanced NILM routinely examines harmonic content up to and including eighth harmonic, and can be customized to examine higher harmonics as necessary.

B. *Transient Detection*

Our advanced NILM recognizes individual loads based on distinctive load transient shapes. The transient behavior of a typical load is intimately related to the physical task that the load performs. For example, the turn-on transients associated with a personal computer and with an incandescent lamp are distinct because charging capacitors in the computer power supply is fundamentally different from heating a lamp filament. Overall transient profiles tend to be preserved even in loads that use active waveshaping or power factor correction. Most loads observed in the field have repeatable transient profiles, or at least sections of the transient profile that are repeatable. Transient-based recognition permits near-real-time identification of load operation, especially turn-on events.

Transients are identified by matching events in the incoming aggregate power stream to previously defined transient signatures, or "exemplars." Exemplars can be determined, for example, by a one-time direct observation of the device in question, or by previous training in the laboratory. Pre-training has proven to be a reasonable approach for very repeatable loads that show up in large quantities, such as fluorescent lamp ballasts. The exemplar may be composed of multiple parts for transients with a number of distinct sections. Each section of the exemplar can be shifted in time and offset to match a incoming transient data. In addition, an overall gain may be applied to all sections of the exemplar to

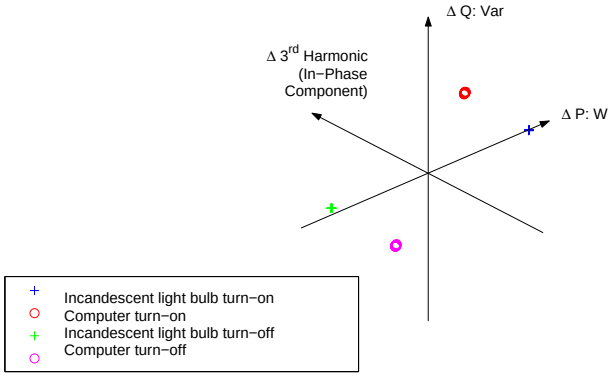


Fig. 2. Computer and incandescent lightbulb turn-on and turn-off event clusters in the ΔP, $\Delta 3^{rd}$ harmonic, ΔQ coordinate system.

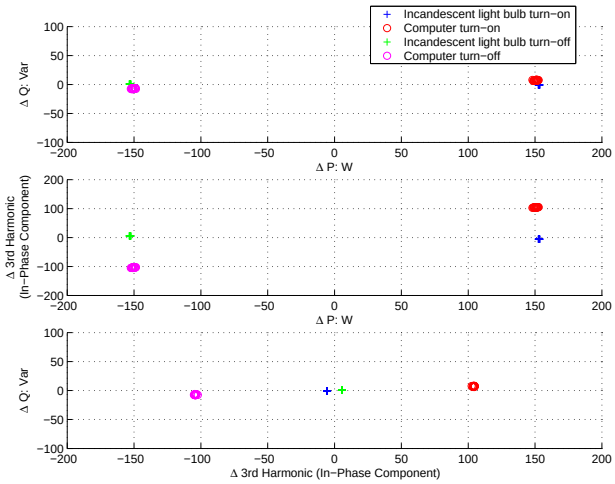


Fig. 3. Three cross-sections of the ΔP, $\Delta 3^{rd}$ harmonic, ΔQ coordinate system, illustrating the separation of clusters.

achieve a better fit. Each event detected is compared to the full set of exemplars by using a least squares criterion to select the appropriate shifts and gains. The match with the lowest residual norm per number of points is then compared to a threshold. If the fit is good enough, the event is classified as a match to the exemplar. If not, the event is left unclassified. Correct classification of overlapping transients is possible using properly designed exemplars.

The value of examining transient information can be understood graphically by examining Figures 4, 5, and 6. Figures 4 and 5 show the spectral envelopes corresponding to the fundamental, in-phase component of current (essentially, "real power") demanded in four different buildings in California. The top graph in Figure 4 shows the power consumed solely by the HVAC panel in a building ("ISD") in Los Angeles County. The bottom trace of Figure 4 and the top trace of Figure 5 show power consumption in two different public schools near San Francisco ("Hanna" and "Pinole", respectively). Finally, the bottom trace in Figure 5 shows the total power consumption in another building in Los Angeles County ("Comm"); the inset in this plot is a magnified portion of the same waveform, illustrating the high density of events. These four traces show a progression in increasing event generation in a building HVAC panel using steady-state signature detection. As the complexity and variation on the electrical grid increases, it becomes more difficult to tune a steady-state change detector to function at all. For buildings with behavior nearing that of the Comm building, it becomes essential to either augment the steady-state approach or discard it completely in favor of transient identification. By enabling the monitor to deal with higher rates of event generation, a single NILM using transient event detection can generally monitor a comparatively more complex building power network than with steady-state change detection alone.

In addition to better load disaggregation, transient analysis can provide diagnostic information. Diagnostics can be developed by exploring the the relationship between the electrical transient and the physics of the load. Specifically, a load model can often be developed which relates the shape of the electrical transient to physical or design parameters of the load. The load model is often a differential equation, although we have experimented with other choices. It is often possible to deduce the physical condition of the load by examining load model parameters when the transient data is rich enough to identify the load model.

Figure 6 illustrates the diagnostic capabilities of our advanced NILM. The numbers displayed in the text window below the graphs in Fig. 6 show estimated parameters of an induction machine model. The solid lines in the graph represent raw data, while the points show the results of a simulation with the estimated parameters. The close agreement shown in the figure indicates that the parameters and model accurately predict the response of the actual load. In addition to developing these physically-based models, we have also developed techniques that can accurately estimate model parameters **without** a good initial guess. These estimation methods enable a building manager to trend load parameters and use parameter information to predict impending faults. In the case of an induction motor, such faults might include shorted motor windings, broken rotor bars, and especially failures in a mechanical system attached to the shaft, e.g., a slipping belt.

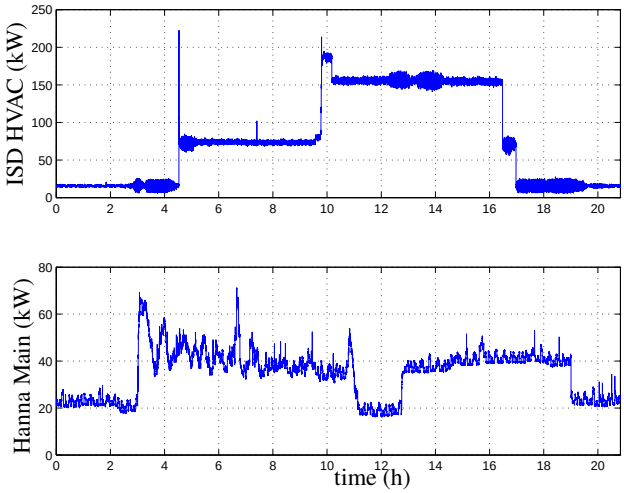


Fig. 4. Plots of the aggregate energy consumption collected at the HVAC service entry in a California county government building and at the main service entrance of a California elementary school.

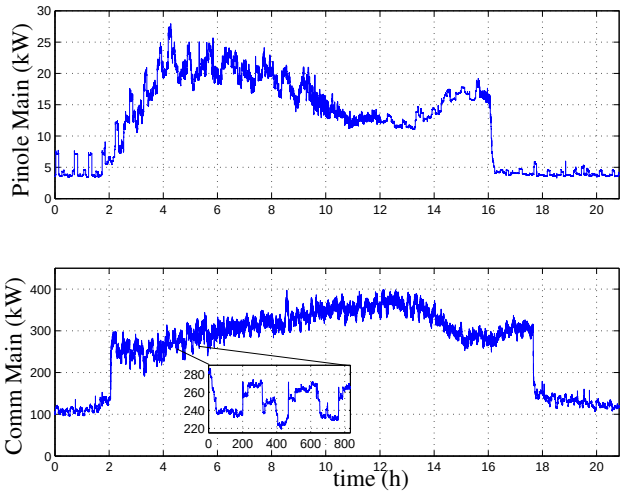


Fig. 5. Plots of the aggregate energy consumption collected at the main service entries of a California middle school and a communications service building. The inset plot illustrates the high density of events; this data in this plot is averaged in order to enhance the visibility of the individual events.

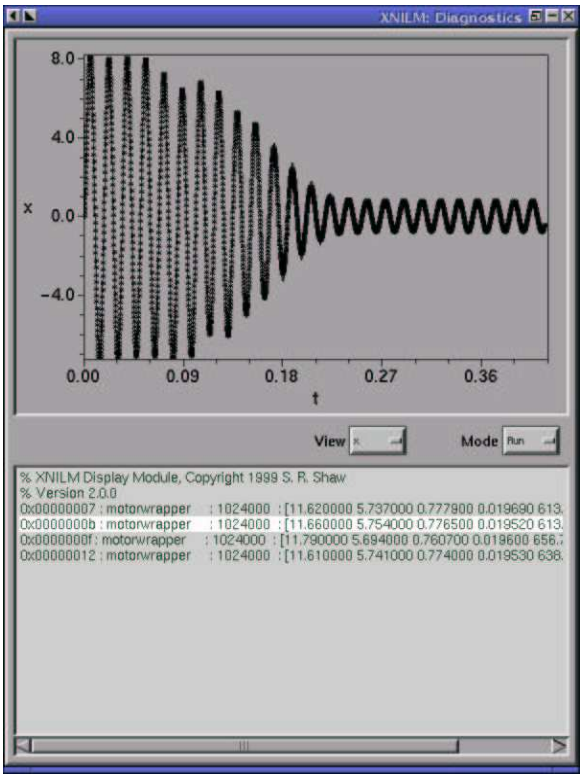


Fig. 6. Startup transient for an induction machine.

C. *Disaggregating Continuously Variable Loads*

A third benefit obtained through transient identification and spectral analysis is that our advanced NILM has the ability to monitor buildings with continuously variable loads. Many such loads, such as the VSDs used in HVAC fans, servo their electrical power consumption under the influence of an active closed-loop controller. Transient detection can simplify the tracking of these and other continuously variable loads in several ways.

Figure 7 shows three spectral envelopes associated with the operation of a 40 Hp VSD in an HVAC plant on the

MIT campus. The top trace in this figure shows the real power demanded by a variable-speed fan drive in an HVAC system. The drive begins with an "open-loop" spin-up to operating speed during the first 40 seconds of operation. This open-loop spin-up is repeatable because a microprocessor controls the startup profile every time the drive is activated. A NILM can recognize that a VSD is active in the building using transient recognition, but transient recognition does not provide a means for continuously tracking the power consumption of a variable load like a VSD.

When the NILM recognizes a continuously variable load like a VSD, we have found that it is often possible to disaggregate the variable load by carefully examining the spectral envelopes. The VSD connects to the utility through a three-phase, delta-connected rectifier. The three-phase rectifier draws distorted, pulsatile current waveforms from each of the three utility phases. As shown in the lower two traces in Figure 7, this rectifier set has the effect of causing the VSD to create characteristic traces not only in real power but also in the fifth and seventh harmonics, as illustrated in the middle and bottom traces respectively. These higher harmonic traces have envelopes that roughly track the shape of the real power trace. If the VSD is the only load at a target site which generates the bulk of a particular harmonic signal (such as the seventh harmonic) in the aggregate data, this particular harmonic signal can be functionally mapped and subtracted from the "real power" spectral envelope in order to remove the continuously varying component of the "real power" signal due to the VSD from the aggregate power signal. This makes it easier to identify the remaining active loads in the aggregate power signal. In general, the higher harmonic information can be used both to track the energy consumption of the variable load and also to disaggregate this variable load from the power consumption of the remaining loads. There are many possible variations on this technique of refining the mappings between a particular harmonic and other harmonics; these techniques allow one group of loads to be disaggregated from another.

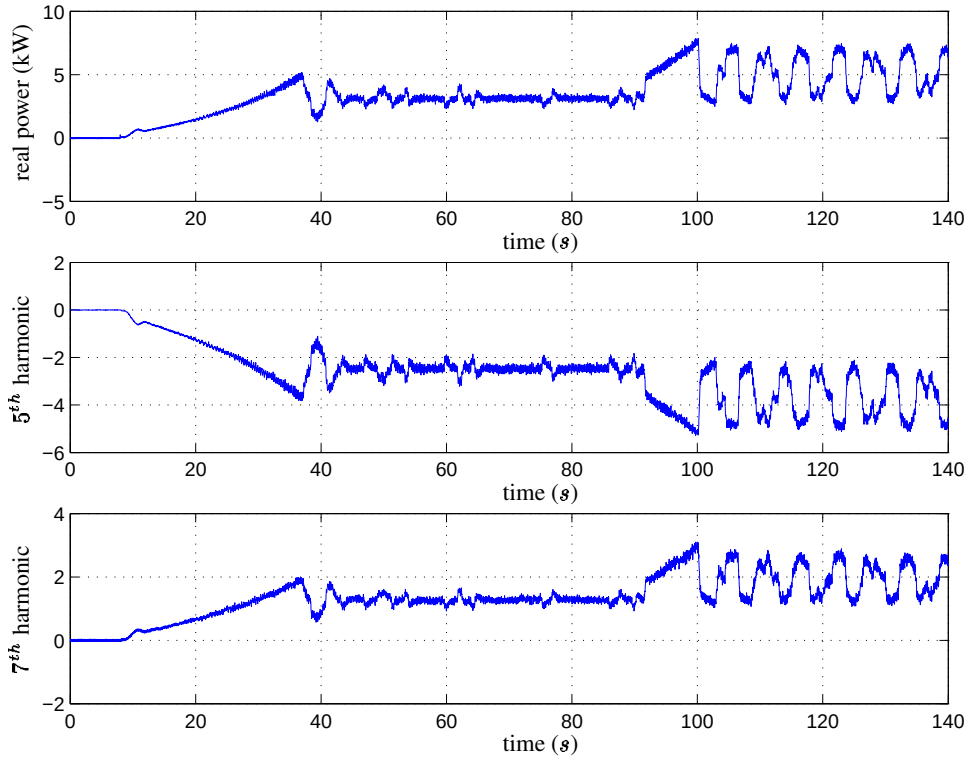


Fig. 7. Startup transient of a variable speed drive.

Additionally, harmonic signatures can provide direct diagnostic information even without parameter estimation. Returning to Figure 7, the drive is operating under closed loop control as it attempts to regulate the pressure in a distant duct by varying fan speed starting at 100 seconds. The control loop is poorly tuned and the drive exhibits a slow, large amplitude oscillation in "steady state." This pathology is not easily caught in a brief inspection of the drive, as the oscillation is fairly slow; nevertheless, the identification of this failure mode is important because it wastes energy and wears the mechanical components of the drive. This condition is easily caught by an advanced NILM which examines transients and spectral envelopes.

IV. Applications and the Future

We have implemented all of these analytical enhancements for nonintrusive monitoring on an inexpensive computer platform for use in the field. A suite of NILM tools, including sophisticated model-based diagnostic algorithms that track (and could in principle trend) model parameters to determine the health of critical loads, has been developed under the Linux operating system environment. All of these software tools run on a Pentium-class personal computer (300 MHz clock or higher) with a PCI-bus data-acquisition card; any relatively inexpensive personal computer, laptop, or embedded system like a PC104 chassis could be used to develop a modern monitor which performs transient event detection. We have used all of these platforms with success in buildings and transportation systems.

Current experiments in the field are aimed at enhancing the diagnostic capabilities of the NILM, integrating it into building energy management and control systems, and improving it as a load monitor for conducting surveys and energy-usage scorekeeping. Even in very modern buildings with sophisticated control and consumption metering systems, we have found that the monitor is invaluable to sophisticated facilities managers. As the NILM requires virtually no additional wiring effort and few installed sensors, it provides invaluable building operating history and health information in a package which is both easy to install and highly reliable.

In the future, we expect to introduce new event detection and classification schemes where load disaggregation and usage tracking are increasingly accurate. The best approaches appear to blend all the tools described above to some degree. For example, the highest accuracy in usage disaggregation may be obtained by combining the steady-state and transient approaches to event recognition. Additionally, the steady-state signature scheme can obviously be enhanced by working in a larger signature space with harmonic data. Yet another combination of steady-state and transient identification methods may be used to alleviate the problem of defining steady-state intervals by using transient event detection to tease apart overlapping transients which occur at high rates. Furthermore, the transient detection approach may be minimally successful when transient shapes are masked or distorted by background, periodic oscillations in the building; by examining changes in the steady-state sections of the aggregate data stream, the transient event detector may be more successful in looking for "most likely" shapes in a cluster of overlapping or distorted transients. We are currently developing "arbiters" that use many different approaches to identify events with the highest possible accuracy and precision.

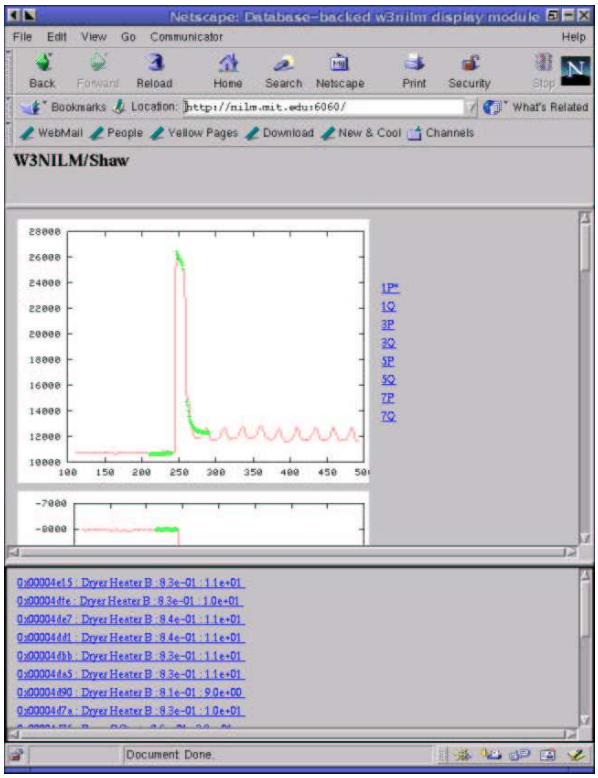


Fig. 8. Nonintrusive monitoring on the web.

Because the advanced NILM uses a conventional computer platform, it is easily connected with communications networks, including the web. An example from a NILM installed on the MIT campus is shown in Figure 8. The Netscape window shows a turn-on transient associated with a washing machine on top of an 11 kW base load; this can be viewed via a remote connection to a computer monitoring a campus dormitory containing a laundromat. The cusp-like steady-state, following the initial acceleration of the agitator motor to operating speed, is due to the "steady-state" oscillation of the agitator. Other load transients recognized by the monitor, including several dryers, are listed in the text window below the graphs. The items in this list are hypertext links corresponding to recorded events; these links can be selected in order to view a record for each indicated transient, including time, energy consumption, and diagnostic parameters (e.g. heater resistance, motor parameters like stator resistance, magnetizing and leakage inductances, etc.) We anticipate that the advanced NILM technology will provide an ideal platform for remote monitoring and for customizing network-accessible information packages for utility customers and building occupants.

ACKNOWLEDGMENTS

This research was funded by the California Energy Commission and a grant from the Grainger Foundation.

REFERENCES

[1] "Wired and Wireless," Technology Review Magazine, Volume 104, No. 5, June 2001, pp. 40–82.

[2] Drenker, S. and A. Kader, "Nonintrusive Monitoring of Electrical Loads," *IEEE Computer Applications in Power*, Volume 12, Issue 4, pp. 47-51.

[3] Hart, G.W., "Nonintrusive Appliance Load Monitoring," *Proceedings of the IEEE*, Volume 80, No. 12, pp. 1870–1891, December 1992.

[4] Leeb, S.B., S.R. Shaw, J.L. Kirtley, Jr., "Transient Event Detection in Spectral Envelope Estimates for Nonintrusive Load Monitoring," *IEEE Transactions on Power Delivery*, Volume 10, Number 3, pp. 1200-1210, July 1995.

[5] Norford, L.K. and S.B. Leeb, "Non-Intrusive Electrical Load Monitoring in Commercial Buildings Based on Steady State and Transient Load-Detection Algorithms," *Energy and Buildings*, Volume 24, pp. 51-64, 1996.

[6] Shaw, S.R., C.B. Abler, R.F. Lepard, D. Luo, S.B. Leeb, and L.K. Norford, "Instrumentation for High Peformance Nonintrusive Electrical Load Monitoring," *ASME Journal of Solar Energy Engineering*, Vol. 120, No. 3, August 1998, pp. 224–229.

[7] Shaw, S.R. and S.B. Leeb, "Identification of Induction Motor Parameters from Transient Stator Current Measurements," *IEEE Transactions on Industrial Electronics*, Volume 46, No. 1, February 1999, pp.139–149.

[8] Shaw, S.R., R.F. Lepard, S.B. Leeb, and C.R. Laughman, "A Power Quality Prediction System," *IEEE Transactions on Industrial Electronics*, Volume 47, No. 3, June 2000, pp. 511–517.

PLACE PHOTO HERE

Christopher R. Laughman received the S.B and the M.Eng degrees from the Massachusetts Institute of Technology in 1999 and 2001, respectively. He is currently a research engineer in the Laboratory for Electromagnetic and Electronic Systems. Current research includes the investigation of modelling and parameter estimation problems in buildings and transportation systems.